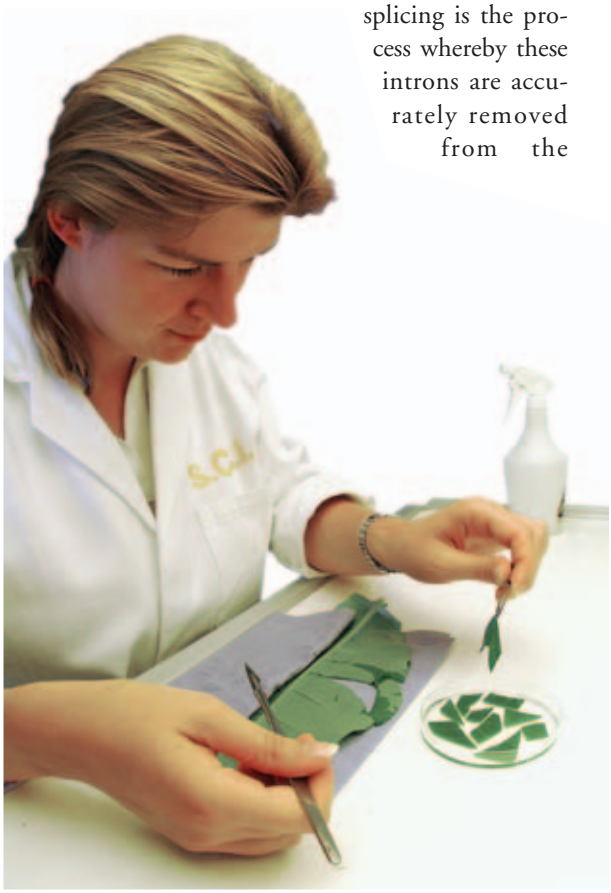
Detailed analysis of two intron splicing signals

C.G. Simpson, G. Thow, G.P. Clark, S.N. Jennings & J.W.S. Brown

Pre-mRNA splicing is one important area where gene expression is regulated. Alternative splicing, where different mRNA transcripts are made from the same transcription unit through alternative splice site choice, is firmly established in humans and *Drosophila* for generating protein diversity and for regulating biochemical pathways. Plants show similar alternative splicing events in many tissue and developmental pathways, and alternative splicing is likely to be extremely important in plant growth and development. To understand how alternative splicing is regulated we need, firstly, to understand the basic mechanisms of splicing in terms of splicing signals and factors. This article provides a detailed analysis of two key plant splicing signals, the branchpoint and polypyrimidine tract, which are recognised to have roles in regulating splice site choice.

A typical plant gene sequence consists of a number of protein coding exons interrupted by non-coding introns. Pre-mRNA splicing is the process whereby these introns are accurately removed from the pre-mRNA to make a mRNA template that is translated into a functional protein. Splicing is an important process, which must be accurate, otherwise non-functional, or dysfunctional proteins can be made. Accuracy of intron removal is achieved by recognition of 5' and 3' splice site signals (Fig. 1) by splicing factors that promote the assembly of the spliceosome. The spliceosome mediates the removal of the introns and the ligation of the exons utilising an intron internal adenosine called the branchpoint found close to the 3' splice site. A consensus sequence surrounding the branchpoint and an associated polypyrimidine tract are firmly established splicing signals in vertebrate introns. They are *cis*-acting signals essential for efficient recognition of the branchpoint nucleotide and for recognition of the downstream splice site. Like the splice site consensus signals, the vertebrate branchpoint consensus is degenerate and poorly conserved (CURAY). The closely associated downstream polypyrimidine tract is a binding site for a protein factor called U2AF^{65} that promotes the selection of the branchpoint.



In plant splicing, these signals have never been properly characterised. We showed previously that mutation of a plant branchpoint adenosine, found in a sequence similar to the vertebrate consensus, significantly reduced splicing efficiency *(Ann. Rep. 1995, 48-49).* Plants do not contain a pronounced polypyrimidine tract as found in vertebrate introns but, have U-rich sequences between the putative branchpoint and the 3' splice site. However, as U- or UA-richness is a feature of efficient plant intron splicing and as U-rich intronic sequences are found throughout plant introns, it has not been possible to distinguish sequences that function as UA-rich elements or as polypyrimidine tracts.

We have been studying the splicing of a 9 nt mini-exon from a potato invertase gene and discovered that mini-exon inclusion is dependent on strong constitutive splicing signals. 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 splice site downstream of the mini-exon. It is likely that these are sites for factors that help define the mini-exon through an exon-defining process that leads to a splicing event that

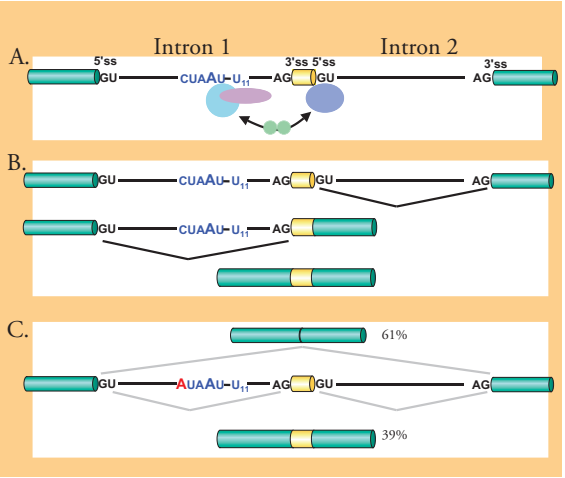


Figure 1 A. Schematic representation of the genomic structure of the invertase mini-exon and the exon definition process that promotes splice site selection of the 9 nt mini-exon. Exons are indicated as boxes and introns by lines. The mini-exon box is highlighted in yellow. The 5' splice sites (GU) and 3' splice sites (AG) are shown. The branchpoint sequence and associated U11 in the upstream intron are shown in blue. Exon definition involves the assembly of factors that bridge the exon. These factors are shown as coloured discs and the bridging interaction shown as a double arrow.
B. The splicing pathway for the invertase mini-exon. In the non-mutated construct (inv1, see Table 1) initiation of splicing removes the downstream intron first, followed by removal of the upstream intron and the formation of the fully spliced product.
C. Disruption of the branchpoint and polypyrimidine tract by single substitution mutation changes the efficiency by which the mini-exon is included in a population of pre-mRNA transcripts and leads to varying levels of mini-exon skipping (lacking the mini-exon box) and mini-exon inclusion (includes the mini-exon box). One change to the branchpoint consensus C→A, highlighted in red, leads to 39% of the pre-mRNA transcripts showing inclusion of the mini-exon and the remaining 61% shows skipping (Table 1). Other single nucleotide substitution changes to these signals are shown in Tables 1 and 2.

removes the downstream intron first followed by the upstream intron *(Ann Rep. 1997/98, 71-73; Ann. Rep. 1999/00, 111-113)* (Fig. 1). The requirement for the branchpoint and U-rich signals for mini-exon inclusion makes this splicing arrangement sensitive to sequencing change leading to skipping of the mini-exon (Fig. 1). The invertase mini-exon system, therefore, provides a system to investigate, in detail, the sequence and spatial requirements of branchpoint and polypyrimidine tract signals for the first time in plants.

A plant branchpoint consensus A systematic mutational analysis of the invertase branchpoint sequence was carried out and the effect on splicing efficiency of the mini-exon in tobacco protoplasts was measured (Table 1). At position -3 from the branchpoint adenosine, a pyrimidine was the optimal nucleotide with a preference for cytidine. A uridine at position -2 was essential as all other nucleotides could not support mini-exon splicing at this position. At position -1 a purine was preferred although pyrimidines could support about 50% splicing of the mini-exon. An adenosine was required as branchpoint nucleotide with a pyrimidine as the nucleotide downstream of the branchpoint. These results clearly support a preference for a consensus sequence which is identical to the vertebrate branchpoint consensus (CURAY) for efficient (>90%) inclusion of the mini-exon. This shows for the first time in plants the importance of the sequence context surrounding the branchpoint nucleotide.

Construct name	Branchpoint mutations					Splicing efficiency %
	-3	-2	-1	0	+1	
inv1	C	U	A	A	U	99
inv74	A	U	A	A	U	39
inv55	G	U	A	A	U	18
inv54	U	U	A	A	U	80
inv56	C	A	A	A	U	3
inv75	C	C	A	A	U	4
inv57	C	G	A	A	U	2
inv73	C	U	C	A	U	51
inv52	C	U	G	A	U	94
inv53	C	U	U	A	U	47
inv76	C	U	A	C	U	7
inv77	C	U	A	G	U	0.4
inv10	C	U	A	U	U	19
inv71	C	U	A	A	A	59
inv51	C	U	A	A	C	99
inv72	C	U	A	A	G	33

Splicing efficiency shows the percentage inclusion of the invertase mini-exon (see Fig. 1C).
Plant branchpoint consensus: CURAY

Table 1 Splicing efficiency of branchpoint mutations.

A plant polypyrimidine tract To examine whether the U-rich sequence downstream of the branchpoint was functioning as a polypyrimidine tract (U and C nucleotides) or a plant UA-rich element (U and A nucleotides), mutant constructs were tested in tobacco protoplasts (Table 2). The constructs maintained a U_4 element and then increasing numbers (2-5) of C or A nucleotides replaced the remaining Us. Splicing to

Construct name	Polypyrimidine tract mutations	Splicing efficiency %
inv1	U_{10}	99
inv38	$U_4A_2U_4$	90
inv41	$U_4A_3U_3$	69
inv58	$U_4A_4U_2$	28
inv60	$U_4A_5U_1$	9
inv39	$U_4C_2U_4$	98
inv42	$U_4C_3U_3$	99
inv59	$U_4C_4U_2$	83
inv61	$U_4C_5U_1$	60
inv42	$U_4C_3U_3$	99
inv64	$U_4C_6U_3$	72
inv65	$U_4C_9U_3$	41
inv8	$U_1C_9U_1$	7

Splicing efficiency shows the percentage inclusion of the invertase mini-exon (see Fig. 1C).
Plant polypyrimidine tract: $U_4Y_{2\text{-}3}U_3$

Table 2 Splicing efficiency of polypyrimidine tract mutations.

include the mini-exon reduced steadily with increasing numbers of A nucleotides such that a construct containing a UA-rich element $U_4A_3U_3$ was splicing the mini-exon below 70% and by $U_4A_5U_1$ splicing was below 10%. On the other hand, C nucleotides were largely able to compensate for U nucleotides and only by the presence of a run of four Cs was splicing of the mini-exon beginning to fall significantly (83%). Therefore, for splicing to include the mini-exon there is a preference for pyrimidines rather than UA-rich elements.

The U-rich region of the invertase gene contains 11 consecutive Us while other invertase genes contain small groups of Us interspersed with other nucleotides. In our previous analyses *(Ann. Rep. 1999/00, 111-113)*, a run of eight Us spliced at 30% and reducing the run further to six and four Us virtually abolished splicing. Although a single U_4 group does not support splicing, as shown here a U_4 group plus a second U_4 or U_3 group can support efficient splicing of the mini-exon. This suggests that a single group of Us is not sufficient to support mini-exon splicing, but a combination of two groups of 3-4 uridines is needed. The separation of these groups was investigated by interspersing them with increasing numbers of Cs. This led to a progressive reduction in splicing efficiency (Table 2), despite the ability of Cs to support splicing in a polypyrimidine tract. Splicing reduced to 41% when a U_4 and U_3 pair was separated by nine Cs, but, this splicing was significantly greater than splicing of a construct containing single U nucleotides separated by nine Cs, which spliced at less than 10%. These results support the requirement for two U-rich groups at least 3-4 nucleotides long and that they should be in close proximity to support efficient splicing. The data suggest that for efficient splicing of the potato mini-exon, the polypyrimidine tract should be $U_4Y_{2\text{-}3}U_3$. This may reflect the binding site of U2AF^{65}, which binds to polypyrimidine tracts and has a binding preference for $U_6C_{2\text{-}3}U_8$.

Using the invertase mini-exon system and an extensive number of mutants we have characterised the sequence surrounding a plant branchpoint and its associated polypyrimidine tract. The identification and characterisation of these signals in plants is significant as these signals are important areas for regulation of splicing and thereby regulation of gene expression. *Trans*-acting proteins mediate 3' splice site selection and, in vertebrate and *Drosophila* systems, compete to regulate alternative 3' splice site choice. It is clear that naturally occurring plant introns show great variation in putative polypyrimidine tracts for splicing and these differences may alter the strength or sequence recognition by plant splicing proteins. It remains to be seen to what extent these plant polypyrimidine tracts may regulate alternative splice site choice and consequently gene expression in plants.