SPLICE SITE SELECTION IN PLANT PRE-mRNA SPlicing

J. W. S. Brown and C. G. Simpson
Department of Cell and Molecular Genetics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, United Kingdom;
e-mail: jbrown@scri.sari.ac.uk; csimps@scri.sari.ac.uk

KEY WORDS: pre-mRNA splicing, intron, exon

ABSTRACT
The purpose of this review is to highlight the unique and common features of splice site selection in plants compared with the better understood yeast and vertebrate systems. A key question in plant splicing is the role of AU sequences and how and at what stage they are involved in spliceosome assembly. Clearly, intronic U- or AU-rich and exonic GC- and AG-rich elements can influence splice site selection and splicing efficiency and are likely to bind proteins. It is becoming clear that splicing of a particular intron depends on a fine balance in the “strength” of the multiple intron signals involved in splice site selection. Individual introns contain varying strengths of signals and what is critical to splicing of one intron may be of less importance to the splicing of another. Thus, small changes to signals may severely disrupt splicing or have little or no effect depending on the overall sequence context of a specific intron/exon organization.

CONTENTS
INTRODUCTION ........................................................... 78
THE SPlicING PROCESS ................................................ 79
PLANT INTRON SEQUENCES ........................................ 80
AU RICHNESS ............................................................. 82
SIMILARITIES WITH OTHER SYSTEMS .......................... 84
3’ Splice Site Selection ................................................. 85
Exon Definition ........................................................ 85
SPECIES DIFFERENCES IN SPlicing ............................... 87

77
INTRODUCTION

Splicing is an integral step in the overall process of gene expression and is one level at which gene expression may be regulated. The excision of intron sequences from precursor nuclear messenger RNA (pre-mRNA) transcripts and ligation of exons is a dynamic yet orderly process that involves the assembly of a large ribonucleoprotein complex, called the spliceosome on pre-mRNAs. The spliceosome is formed by the sequential assembly of small nuclear ribonucleoprotein particles (snRNPs) and other non-snRNP spliceosomal proteins. It catalyzes the precise recognition and cleavage of the intron and subsequent ligation of the exons. Failure to remove introns or any inaccuracy in intron excision would alter the open reading frame of the mRNA leading to altered protein size, sequence, and probably function. Inaccurate splicing may affect mRNA stability or transport. The accuracy of splicing is mediated by conserved sequences in the pre-mRNA and by the spliceosomal snRNPs and other factors that recognize these sequences (reviewed by 5, 46, 47, 71).

The availability of in vitro splicing extracts from animal and yeast cells along with yeast RNA processing mutants has facilitated the dissection of the splicing process in these organisms. The lack of an in vitro splicing extract for plants has meant that investigations into plant pre-mRNA splicing have progressed at a much slower pace. Nevertheless, splicing in plants is particularly interesting. First, there are clear differences in plant splicing compared with fungi and animals. Second, splicing differences exist between monocotyledonous (monocot) and dicotyledonous (dicot) plants (29, 41). More recently, examples of splicing differences between different species of dicots have been found, suggesting yet another level of subtlety in plant intron signal recognition (39, 65). Third, differentiation in animals often requires alternative or differential splicing to produce cell-type specific mRNAs from a single transcript; this selection of alternative splice sites produces proteins with different activities (89). A number of examples of alternative splicing have been described in plants, but currently little information is available on the functional significance of the alternative products. As more genes are analyzed, we anticipate more examples. If alternative splicing is an important factor in the control of plant development and differentiation, an understanding of intron sequence and structure, the factors involved in splice site recognition, and spliceosome assembly will be a prerequisite to studying the regulation of differential splicing.

Over the past 10 years, various aspects of plant intron structure, spliceosomal components, and splicing have been investigated. Much emphasis was placed
initially on the differences between plant and animal introns and their splicing behavior in reciprocal assays. In particular, the main distinguishing feature of plant introns is AU and U richness in relation to flanking exons, a feature which has received a great deal of attention. The first decade of analysis of plant splicing has been reviewed thoroughly by Simpson & Filipowicz (85).

Because of their comprehensive review, we do not wish to reiterate large areas of research. Although the review was published relatively recently, a number of new and interesting observations have been made, allowing current hypotheses of the mechanism of splice site selection to be reassessed. In particular, we discuss the changing perceptions of the role of various splicing signals in splice site selection, the importance of exon sequences (particularly with regard to exon scanning in plant intron splicing), and how these signals may interact with the important AU-rich sequences in the intron. Other advances have been made in the areas of alternative splicing and splicing regulation, communication between introns, and in the discovery of a novel second class of pre-mRNA introns with (normally) noncanonical 5' and 3' splice sites.

THE SPLICING PROCESS

Spliceosome assembly and the spliceosome cycle have been determined mainly using vertebrate and yeast systems. Introns are excised in a two-step cleavage-ligation reaction, where the first step involves cleavage at the 5' splice site with formation of an intron lariat at an adenosine nucleotide (the branchpoint), usually 18–40 nucleotides (nt) upstream of the 3' splice site. In the second step, following cleavage at the 3' splice site, the exons are ligated, and the intron is released as a lariat, which is then debranched and degraded (46, 47, 71).

The accuracy of splicing depends on the ability of the splicing machinery to recognize various intron signals. This recognition is mediated by spliceosomal U-type small nuclear ribonucleoprotein particles (snRNPs) and numerous non-snRNP protein factors such as hnRNP proteins (reviewed in 20, 42), SR proteins (a family of proteins with arginine-serine-rich domains) (reviewed in 22, 61), and DEAD- or DEAH-box–containing proteins (RNA-dependent ATPases or ATP-dependent RNA helicases) (reviewed in 5, 23). HnRNP and SR proteins both contain RNA-binding domains and auxiliary domains, often glycine-rich in hnRNP proteins, and serine-arginine-rich in SR proteins, which are involved in protein-protein interactions (22, 42). The DEAD- and DEAH-box proteins are thought to mediate the conformational changes that occur in spliceosome assembly and disassembly (14).

The spliceosome cycle has been reviewed in detail (5, 46, 47, 71). Briefly, the pre-mRNA transcript becomes associated with hnRNP proteins, which may define intron/exon regions by virtue of sequence-specific binding and aid the
association and stability of other splicing factors. U1snRNP is recruited to the 5′ splice site where the 5′ end of U1snRNA interacts with the 5′ splice site through base pairing. The polypyrimidine tract (lying between the branchpoint and 3′ splice site) is bound by U2AF65 (vertebrates)/Mud2p (yeast) and interacts with the U1-5′ splice site complex, through cross-intron bridging interactions, bringing the 5′ and 3′ splice sites into close proximity. In yeast, the branchpoint binding protein (BBP) interacts specifically with the branchpoint and with the proteins, Mud2p at the polypyrimidine tract, and Prp40, a component of the U1snRNP (1, 7). This commitment complex is the earliest functional intermediate in spliceosome assembly and targets the pre-mRNA to the splicing pathway. U2AF65 and ATP are required for the association of U2snRNP, which also involves base pairing between the branchpoint sequence and U2snRNA, resulting in bulging of the branchpoint nucleotide, usually an adenosine (78). Formation of the presplicing complex requires a number of splicing factors, including SR proteins, that may function in promoting or stabilizing interactions between components (5, 46, 47). The addition of the U4/U6-U5 tri-snRNP forms the splicing complex in which a number of conformational changes occur to bring the interactive splice sites together. These changes require ATP and splicing factors such as RNA helicases/ATPases and involve a series of interactions between the pre-mRNA, snRNAs, and protein components (5, 46, 47, 60, 71). Ultimately, the reactive sites of the pre-mRNA are arranged into the correct conformation [catalytic center(s)] in which the splicing reactions occur. Following catalysis, the spliced exons are released, the postsplicing complex containing U2, U4, U5, and U6snRNPs is disassembled and the snRNPs recycled, and the intron lariat debranched and degraded.

PLANT INTRON SEQUENCES

Plant intron 5′ and 3′ splice site consensus sequences are very similar to those of vertebrate introns, although individual introns exhibit great variation around the highly conserved :GU and AG: dinucleotides (11, 13, 15, 21, 33, 35, 55, 58, 85, 86; Figure 1). In yeast, the internal branchpoint sequence is absolutely conserved—UACUAAC, where the underlined adenosine is the branchpoint nucleotide. In animals, the branchpoint sequence is variable, with the consensus YURAY (45), usually positioned 18–40 nt upstream of the 3′ splice site. Only recently have branchpoints been mapped in plant introns and shown to be an important splicing signal (49, 83). Again, the plant intron branchpoint consensus is very similar to that of vertebrates (49, 83, 99). In vertebrate introns, the fourth splicing signal required for efficient splicing is a polypyrimidine tract, found directly upstream of the 3′ splice site. In some yeast introns, a poly U sequence in this region is necessary for splicing. In plant introns, this region is often
U-rich and potentially may be involved in branchpoint definition and 3′ splice site selection as in vertebrate introns (4).

The major difference between plant introns and those of vertebrates and yeast is that plant, and particularly dicot, introns are AU rich. In fact, in an analysis of base composition of 271 dicot and 146 monocot plant introns in the regions extending 50 nucleotides upstream and downstream of the 5′ and 3′ splice sites, the A content of introns matches that of the flanking exons. Introns are ~15% more U-rich than the flanking exons, which are ~15% more GC-rich (85). This property and the associated difference in nucleotide composition between introns and exons are important factors in pre-mRNA splicing in plants (28, 29, 105). The emphasis of many splicing studies in both dicot and monocot systems has been aimed at defining the role of AU richness in general, and at identifying the functional elements of this characteristic, such as U- and AU-rich elements and AU/GC intron/exon border sequences in particular, and is discussed in detail below (4, 15, 16, 26, 28, 29, 53–55, 57, 58, 66–68, 82, 85; Figure 1).

Recently, the AT-AC introns were identified in vertebrates as a minor class of introns with nonconventional splice site sequences (31, 107). AT-AC introns

![Diagram](image.png)

*Figure 1* Model of early spliceosomal complex formation in plants. (A) Important intron signals: GU, 5′ splice site; AG, 3′ splice site; A, branchpoint; UA, U-rich elements; U, U-rich elements between the branchpoint and 3′ splice site and GC/AG-rich exonic elements. (B) Exon definition: By analogy to assembly of the vertebrate commitment complex, U1snRNP is recruited to the 5′ splice site. In plants, this may be directed or the interaction may be stabilized by factors that bind to UA-rich elements in the introns. At the 3′ splice site/branchpoint, factors presumably similar to factors such as U2AF65 assemble in the branchpoint region. This may also be mediated by factors binding to UA- or U-rich elements which aid definition of the branchpoint and 3′ splice site. Once factors are assembled and stabilized across an exon, interactions between the 5′ splice site and branchpoint region across the intron will form the commitment complex before U2snRNP addition and functional spliceosome assembly.
have an invariant :AUAUCCUY sequence at the 5′ splice site, YAC: at the 3′ splice site, and a presumptive branchpoint (UCCUURAY) 16–19 nt upstream of the 3′ splice site (31, 107). The mechanism of splicing of these introns is a two-step transesterification reaction, as for conventional introns, but recognition of the splicing signals requires a unique set of snRNAs (32, 95, 96). U11 and U12 basepair with the 5′ splice site and branchpoint of AT-AC introns respectively in an analogous manner to U1 and U2 of the major class of spliceosomes (32, 43, 95). U5 is common to both spliceosomes (95), and U4atac/U6atac snRNAs are found as a di-snRNP and are essential to AT-AC spliceosome function (96).

In plants, two examples of AT-AC introns have been described: intron 14 of the RecA-like protein gene and intron 7 of the G5 gene from Arabidopsis are AT-AC introns, although the latter contains an AA: dinucleotide at the 3′ splice site (106). Other examples of such introns have been proposed based on computer analysis. A second class of AT-AC introns with canonical splice sites (GU-AG) but with other AT-AC intron features (e.g. the proximity of the branchpoint to the 3′ splice site) have also been proposed (P Rouzé, personal communication). Whether these introns have a function in regulating levels of their host mRNA remains an important question.

AU RICHNESS

Splicing in plants has progressed, for the most part, using transient expression analysis. Plasmid constructs consisting of a constitutive promoter, the test intron with complete or partial flanking exon sequences, and/or intron-containing reporter genes are introduced into plant protoplasts. The accuracy and efficiency of splicing is determined by Northern, RNase A/T1 protection mapping; RT-PCR; or reporter gene expression. While an analysis of splicing in vivo rather than in vitro may well be an advantage for assessing true splice site choice and efficiency, there are limitations that need to be considered. First, strong constitutive promoters designed to give high levels of the transcript of interest can mean abnormal levels of pre-mRNA transcript, which may overload the splicing process. Second, most analyses to date of plant splicing use single intron constructs, taken out of their authentic gene context, and often tested in a heterologous species. There is evidence that introns are influenced by the presence of their neighboring intron and exon sequences (12, 15, 28, 66) and that splicing variation exists among plant species (29, 39, 41, 65). These limitations have the potential to give information that may not be a precise representation of splicing in a normal genetic background. Despite these weaknesses, studies so far have established the key principles for splicing in plants. By analyzing the increasing number of splicing mutants, especially in Arabidopsis, many of
these limitations are removed. This, along with the ability to prepare transgenic plants, will permit us to test these principles and establish a more detailed understanding of plant splicing processes and mechanisms.

The differences between plant and animal splicing systems were highlighted by variable splicing of animal or plant/animal hybrid introns in plant cells. For example, the only animal introns that are efficiently spliced in plants to date are the human β-globin intron 1 assayed in maize (29), the SV40 small-t intron (37), and the mouse hsc70 intron 5 (JWS Brown, unpublished data) assayed in tobacco. These animal introns are all AU rich.

Goodall & Filipowicz (28) first demonstrated the requirement for AU richness for efficient plant intron splicing using synthetic introns assayed in tobacco cells. They proposed roles for the AU sequences either as binding sites for putative AU-binding proteins or sites to reduce secondary structure formation. Of great influence in the formulation of models for plant intron splicing were the findings that intron splice sites bordering AU-rich sequences were accurately and efficiently used despite the lack of a conserved branchpoint or polypyrimidine tract (28). The suggestion that a conserved branchpoint was not required and that presumably any adenosine (or other nucleotide) in the AU-rich intron could be used led to the idea of plant introns being defined by virtue of binding of specific proteins to AU-rich elements within the intron (28). This model was developed further with a series of experiments using pea and maize introns in *Nicotiana benthamiana* cells that suggested that 5′ and 3′ splice sites were selected because of their proximity to the transitions between the AU-rich intron and AU-poor exon. In these experiments, the positions of AU/GC borders were moved relative to splice sites by insertion and deletion of AU-rich and GC-rich sequences and mutation of authentic splice sites (53, 54, 67). The data suggested activation of splice sites at AU/GC transitions and “masking” of splice sites embedded in AU-rich regions. Similar insertional experiments with monocot introns tested in maize cell transient assays also supported the transition model, but it was recognized that in some cases an “internal” signal appeared to be needed (58). Splicing of nonintronic AU-rich regions (82) also supported the hypothesis that the only requirements for efficient plant intron splicing were AU-richness with adjacent splice sites (28, 53, 54, 58, 67). The importance of plant exon sequences was demonstrated by analyzing splicing of introns where GC- and AG-rich elements were inserted or were present in flanking exons, altering splicing efficiencies (15, 68; Figure 1). The key elements of AU-rich intron sequences were first defined as AU islands of 4–7 nt and later as U-rich elements with the ability to bind nuclear proteins (26, 53–55, 67). It is assumed that U-binding proteins associate with intron sequences and presumably a different set of proteins with different specificity bind to exon sequences, thereby aiding the delimitation of the exon/intron (21, 28, 29, 53–55, 67, 68, 85).
Despite the weight of experimental evidence, some observations of plant intron sequence and splicing behavior are not entirely consistent with intron definition on the basis of AU sequences and AU/GC transitions. First, while consensus plots of AU/GC content highlight intron/exon borders (85, 105), as with all consensus sequences, there is a great deal of variation on the individual gene sequence level. Some plant introns do not exhibit a significant or, in some cases, any AU/GC differential, and some exons contain AU islands in close proximity to splice sites. For example, a comparison of neighboring exons and introns in 209 Arabidopsis genes found 10% of the genes to contain at least one exon/intron transition where the exon has between 0% and 8% higher AU content than the intron (CG Simpson & JWS Brown, unpublished data) that, if bound by proteins, might be expected to interfere with splicing. In addition, splice sites found within AU-rich regions or GC-rich regions can be selected. An authentic splice site was surrounded by GC-rich sequences and utilized 85% of the time (77), and a splice site embedded in AU-rich sequence was efficiently selected when a competing distal 5′ splice site was mutated (68). On the other hand, when dicot/monocot (AU-rich/AU-poor) hybrid introns were introduced into dicot cells, the authentic monocot intron splice sites were efficiently selected despite being some distance from the AU/GC transition formed by the intron fusions, and potential splice site–like sequences more proximal to the transitions were not selected (83). Thus, local AU/GC transitions around splice sites may not be major determining signals but reflect the method by which plant equivalents of hnRNP proteins exert their sequence specificity.

SIMILARITIES WITH OTHER SYSTEMS

The similarities among eukaryotes are highlighted by the strong conservation of splicing signals, snRNAs in important regions of primary sequence or secondary structure (91), the ever-increasing number of plant homologues to yeast and animal spliceosomal proteins (27, 79, 84, 87; JWS Brown & CG Simpson, unpublished data), SR proteins (48, 51, 52) (in particular, homologs of U2AF65; C Doman & W Filipowicz, personal communication), a putative hnRNP C1/C2 (J Turner, personal communication), and the evidence for similar mechanisms of splice site selection operating in plants (see below). At the whole cell or nucleus level, there is also extensive organizational conservation with nuclear spliceosomal components being distributed throughout the fibrous interchromatin nuclear network and in coiled bodies directly paralleling that seen in animals (8, 18, 25, 85). Altogether, these similarities suggest that fundamental differences in splicing between plants and animals lie at the level of intron sequence recognition early in spliceosome formation. It seems likely that the distinguishing features of plant pre-mRNAs and proteins specific to them
function in establishing the commitment or presplicing complexes after which spliceosome assembly and splicing will be very similar to other systems.

3′ Splice Site Selection
Mutation of the conserved :GU and AG: dinucleotides abolishes correct splicing and optimizing splice site sequences can increase splicing efficiency (16, 26, 29, 83). The most important aspects of splice site selection in terms of this review are the similarities of the splicing behavior of many plant introns to that of vertebrate introns. For example, in mammalian intron splicing, the scanning model for 3′ splice site selection proposes that the spliceosome scans from the branchpoint and usually selects the first AG: dinucleotide downstream of the branchpoint (90). If more than one AG: lies in close proximity, local scanning in this region may select the best 3′ splice site on the basis of sequence context (88). In addition, when branchpoint, 3′ splice site and associated exon fragments are found in duplicate, the downstream-most splice site was generally selected (80). There is strong evidence that such selection mechanisms operate in plant splicing (12, 83 and references therein). From experiments with a 3′ splice site sequestered in a stem-loop structure, early recognition of the 3′ splice site in plant spliceosome assembly was demonstrated and was suggested to be involved in commitment complex formation or branchpoint selection (50). These examples parallel yeast and vertebrate splicing where the 3′ splice site is recognized early in spliceosomal assembly and aids identification of the internal branchpoint sequence. This is further supported by the need for a conserved branchpoint sequence for efficient splicing (83). Early recognition of the 3′ splice site may involve the positioning of AU-binding proteins at the exon/intron border and subsequent recruitment of splicing factors analogous to vertebrate factors (e.g. U2AF) that recruit the U2snRNP to the branchpoint allowing formation of a commitment complex. In terms of exon definition (see below), interactions between factors at a 3′ splice site and factors defining the preferred branchpoint sequence may be a key process in correct 3′ splice site selection. This model, which parallels the vertebrate and yeast models, would allow accurate selection of 3′ splice sites for introns that lack a AU/GC transition and are located within AU-rich or GC-rich regions (Figure 1).

Exon Definition
Exon sequences have been known to be important in vertebrate intron splicing for many years. In mammals exons tend to be small (<300 nt) and introns very large. A model for exon definition or scanning in splice site selection has been proposed whereby interactions between factors at a 3′ splice site and a downstream 5′ splice site define the exon (reviewed in 6, 9). In some cases of regulated splicing, interactions can occur between 5′ or 3′ splice site factors and
proteins bound to sequences in an adjacent intron (intron enhancers) or exon (exon enhancers) to define the splice site (100). For example, stable U1snRNP binding to a 5′ splice site can promote U2AF binding to an upstream 3′ splice site (36), and binding of SR proteins or U1snRNP to purine-rich exon enhancer sequences enhances splicing of the upstream intron (98, 101). Mutation in splice site sequences can disrupt these interactions, often leading to exon skipping (removal of both introns and the intervening exon from the pre-mRNA). In contrast, lower eukaryotes with usually short introns may not require exon definition, and factors may interact directly across intron sequences (intron definition; 6).

In plants, although exon sequences have long been recognized as important in splicing by virtue of the contrast in base composition to adjacent introns, only recently have more active roles been discovered. Analysis of mutants in various Arabidopsis genes, with mutations in and around splice sites, has uncovered a number of examples of exon skipping in plant splicing (12). To date, six such mutations have been described: ag-4, cop1-1, cop1-2, cop1-8, spy-1, and spy-2 (38, 70, 108; CG Simpson & JWS Brown, manuscript in preparation; Figure 2).

Figure 2  Exon definition in plant splicing. In multi-intron transcripts, factors assembled across an exon aid selection of the splice sites. Mutations to either the 5′ or 3′ splice sites flanking the central exon will disrupt assembly of the splicing components at these sites leading to a loss of interaction/stabilization across the exon and exon skipping.
The two mutant alleles of SPINDLY, spy-1 and spy-2, both give rise to skipping of exon 8 (38). The spy-2 allele contains a G → A mutation at the 3′ splice site -AG of intron 7, and the spy-1 mutation is to the last nucleotide in exon 8. The cop1-1 and cop1-2 mutants cause exon skipping of exon 6. Cop1-1 has a mutation in the 3′ splice site of the upstream intron, cop1-2 is mutated in the 5′ splice site of the downstream intron, and cop1-8 carries a mutation in the 3′ splice site of intron 10, which leads to removal of introns 10 and 11 and exon 11 (CG Simpson & JWS Brown, unpublished data). Finally, in the ag-4 mutant, the majority of transcripts undergo exon skipping of exon 6 (108). In the exon skipping mutants, it would be expected that splicing factors would associate with the adjacent unmutated splice site, but the lack of, or unstable binding of, proteins to the mutated splice site would impair cross exon interactions and block correct splicing.

The examples of exon skipping in the *Arabidopsis* intron mutants provide strong evidence for exon definition (12; Figure 1). If the intron is the unit of definition in plant splicing, mutations to splice sites would not be expected to affect splicing of neighboring introns (6, 9). Therefore, splicing of at least some plant introns may involve exon definition and require direct interactions across the intron to form the spliceosome. One implication of exon definition in plant splicing is that splicing of individual introns in a multi-intron pre-mRNA may be affected by other introns in the transcript. Such cooperation or communication could occur by protein interactions where factors on one intron promote factor assembly at other introns. Evidence for such cooperation among vertebrate introns comes mainly from human genetic mutations where mutation of one intron can affect processing of adjacent or even distal introns (2, 76) and by in vivo splicing analyses where the presence of one intron is required for maximal splicing efficiency of other introns (72, 73).

**SPECIES DIFFERENCES IN SPICING**

Dicots need ~60% AU for efficient splicing, while monocots are more flexible and can splice introns with as little as 30% AU. This difference between the AU-content requirement for splicing in monocots and dicots may reflect poorer affinity of the dicot AU-binding proteins for introns with few or only short AU-rich sequences. The apparently more flexible monocot splicing machinery would suggest that monocot hnRNP protein analogs have a broader sequence specificity that is reflected in the range of AU content seen in monocot introns (29). Besides the monocot/dicot difference, differences in splicing behavior of pre-mRNA transcripts among different species are beginning to emerge. For example, the relative usage of combinations of different splice sites in the Ds transposable element of maize differs between tobacco and *Arabidopsis* (75).
Similarly, cryptic splicing of the fourth intron of the Ac transposase gene in *Arabidopsis* has been postulated as the reason for poor levels of Ac excision in transgenic *Arabidopsis* (39, 65), compared with other dicotyledonous plants, such as tomato and tobacco, and in its host plant, maize (3, 40). Finally, expression of the *Aequorea victoria* GFP gene was severely hampered by a cryptic splicing event that removed an 84-nt intron with the GFP coding region (34). The cryptic intron (68% AU) contained 5′ and 3′ splice site and branchpoint sequences with a good match to plant intron consensus sequences. In *Arabidopsis*, this intron was efficiently spliced in almost 100% of transcripts, giving little evidence of fluorescence in transgenic plants. However, in tobacco, splicing efficiency of the cryptic intron was 40%, resulting in expression of the protein. Thus, considerable variation exists in the ability of the splicing apparatus of different dicot species to recognize and excise the same intron sequence.

**ALTERNATIVE SPLICING**

In animals, mechanisms exist to regulate splice site choice via alternative splicing, which can result in either retention or inclusion of introns in the open reading frame; selection of alternative 5′ and 3′ splice sites, which leads to increases or decreases in the size of particular exons; selective inclusion of mutually exclusive exons; and exon skipping (69, 89, 98). Alternative splicing results in the production of different mRNAs that encode proteins with functional differences often in a tissue-specific or developmental stage-specific manner and is therefore an important regulatory mechanism in gene expression.

In plants, a number of examples of alternative splicing have emerged, but little functional or developmental information is available. A common form of alternative splicing in plants is intron retention (27, 74, 85, 93, 97), which probably reflects poor recognition of the intron rather than active processes inhibiting the splicing reaction. The presence of a poorly spliced intron in a pre-mRNA can potentially regulate transcript levels because of competition with other RNA-processing events such as polyadenylation. Polyadenylation sites present within such an intron may compete with the splicing apparatus with the result that many transcripts are prematurely polyadenylated (39, 59, 85). In vertebrates, the presence of in-frame premature stop codons within intron sequences can lead to effects on mRNA stability, transport from the nucleus, translation, and subsequently levels of gene expression (17, 62). Alternative splicing of the *wxG* allele (G retrotransposon insertion in *waxy*) in maize shows a down-regulation of expression in endosperm but not in pollen. This tissue-specific effect may be related to the presence of premature termination codons within exon 13 (63). Alternatively, levels of *Arabidopsis* U1 70k mRNA transcripts that retain intron
7 showed variable levels of expression in different tissues, but in this case there is no premature termination codons or polyadenylation within the intron (27), suggesting that other mechanisms linked to splicing may operate to alter levels of gene expression.

Selection of alternative 5′ and 3′ splice sites occurs in a number of plant introns. In some cases, both splice sites are selected in a similar ratio, such as in rubisco activase from a number of species (103). In others, the relative usage displays different cell- or tissue-specific splicing patterns. For example, alternative 5′ splice site selection in the third intron of chorismate synthase (LeCS2) from tomato and alternative 3′ splice site selection in the first intron of the H-protein subunit of glycine decarboxylase from Flaveria trinervia show an alteration toward more distal and weaker splice sites in floral and root tissue, while stronger proximal sites are predominantly selected in leaf, stem, and cotyledon (30, 44). In animal systems, particular splicing factors can alter the splice site choice of competing splice sites (109), such that the tissue or organ-specific splicing pattern differences may reflect different splicing factor levels. Similarly, in animals, splicing patterns can be altered in response to stress; however, in plants, splicing patterns tend to be unaffected by stress situations (64, 85). Recently, however, an example of exon skipping has been described in which a mini-exon is skipped at low levels in response to cold storage conditions (10). This effect may reflect lower splicing accuracy or changes in splicing factors under these physiological conditions, but because the mini-exon encodes part of the most conserved region of the protein, the possibility remains that the exon skip produces a functionally modified peptide.

Although the functional significance of alternative splicing systems in plants is mostly unknown, the effect of alternative splicing events can be extensive on gene expression. For example, intron retention can lead to alterations in cellular location for functional peptides when they are found separating a signal sequence from the functional peptide. Use of a translation start codon in the intron can produce cytosolic forms of otherwise targeted proteins (19, 81, 97). The different forms of alternative splicing can also lead to the production of truncated peptides due to premature termination either by activation of intronic polyadenylation signals or by bringing translation stops in frame (24, 59, 92, 93). The role of these truncated peptides, with the exception of secreted and membrane-bound SLG (94), is not known, but in some cases significant levels of the truncated peptide accumulate (59). It is feasible that such peptides are involved in regulation, possibly acting as trans-dominant inhibitors. Alternatively, competition between processing signals such as splicing and polyadenylation could regulate the levels at which the functional peptide is found. Overexpression of the flowering control gene, FCA, resulted in only a limited acceleration of
flowering time, reflected by only a small increase in fully spliced mRNA transcript. However, the levels of the short polyadenylated transcript, which terminates in intron 3, showed a 150-fold increase over the wild type. Thus, the default pathway for polyadenylation in intron 3 may keep the levels of FCA at the correct level to keep flowering time constant (59). Finally, it has long been known that the insertion of transposable elements into gene sequences can result in gross changes in the processing of the pre-mRNA transcript (55, 56, 102, 104). Often the transposable element inserts into intron sequences and is then removed by splicing (using cryptic splice sites within the transposable element) but often inaccurately such that pre-mRNA encodes a protein that differs from the wild type and often has reduced functionality. Although the number of alternative splicing systems in plants is increasing, the importance of this area of gene regulation is still in its infancy, and a better understanding of the functions of different peptides and the underlying splicing mechanisms is readily awaited.

CONCLUDING REMARKS

As more information on animal, yeast, and plant splicing and their splicing components accrues, it is becoming clear that greater similarity exists than was previously thought. It is, for example, inconceivable that a protein such as PRP8, known to be essential in animal and yeast systems for many stages of spliceosome assembly and which is so highly conserved in plants, should not carry out the same functions in plant splicing. The recent demonstrations of the importance of branchpoint sequences and exons in plant splicing again draw closer parallels with other eukaryotic systems in terms of mechanisms of splice site selection. This degree of similarity would suggest that the distinguishing feature of UA-rich sequences in plant introns appears to be involved early in intron recognition and prespliceosomal complex formation. Its particular role and the factors that recognize this intron signal still remain to be resolved fully. The increasing number of alternatively spliced gene systems, the characterization of splicing mutants, the examples of exon skipping and recent experiments on test introns all underline the subtle complexity of the splicing process and how splicing of any particular intron depends on the balance of a number of signals and factors. An appreciation of this balance will be necessary in understanding how gene expression can be regulated at the level of pre-mRNA splicing. The study of splicing of specific gene systems in their own context may provide new insights into our knowledge of plant splicing.
SPLICING IN PLANTS

28. Goodall GJ, Filipowicz W. 1989. The AU-rich sequences present in the introns of...
plant nuclear pre-mRNAs are required for splicing. Cell 58:473–83
41. Keith B, Chua NH. 1986. Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. EMBO J. 5:2419–25
46. Krämer A. 1995. The biochemistry of pre-mRNA splicing. See Ref. 47a, 3:35–64
SPLICING IN PLANTS


88. Smith CWJ, Chu TT, Nadal-Ginard B. 1997. The ex-


years old. *Nat. Genet.* 14:383–84

