Pre-mRNA splicing in higher plants

Zdravko J. Lorković, Dominika A. Wieczorek Kirk, Mark H.L. Lambermon and Witold Filipowicz

Most plant mRNAs are synthesized as precursors containing one or more intervening sequences (introns) that are removed during the process of splicing. The basic mechanism of spliceosome assembly and intron excision is similar in all eukaryotes. However, the recognition of introns in plants has some unique features, which distinguishes it from the reactions in vertebrates and yeast. Recent progress has occurred in characterizing the splicing signals in plant pre-mRNAs, in identifying the mutants affected in splicing and in discovering new examples of alternatively spliced mRNAs. In combination with information provided by the Arabidopsis genome-sequencing project, these studies are contributing to a better understanding of the splicing process and its role in the regulation of gene expression in plants.

Properties of plant introns

The intron and exon organization of higher plant genes is similar to that of vertebrates. Most plant genes (80–85%) are interrupted by introns, and a single gene might contain >40. Plant introns are generally shorter than those in vertebrates: about two-thirds are <150 nt long, ranging in size from ~60 to 10 000 nt. The consensus sequences of the 5’ ss and 3’ ss, AGUAGAG and TGCCAG, respectively, are similar to those in vertebrates (Fig. 2). As in other organisms, plant pre-mRNA splicing is a two-step process, involving two trans-esterification reactions and lariat formation. The branch point region, with its loose consensus and the ~30 nt distance from the acceptor AG, resembles that of vertebrate introns, but clearly differs from the highly conserved branch point of yeast (Fig. 2). However, the positions of only two plant branch points have been determined experimentally.

In spite of these similarities, the requirements for intron recognition in plants differ from those in other eukaryotes, and plant cells generally fail to splice heterologous pre-mRNAs. The most important difference is a strong compositional bias for UA- or U-rich sequences in plant introns compared with those from yeast and vertebrates (Fig. 2). On average, in both dicots and monocots, introns are ~15% more UA-rich than exons, the U residues being largely responsible for this difference. U-rich sequences, typically distributed throughout the entire length of plant introns, are required for efficient intron processing and splice-site selection; this position-independent function distinguishes them from vertebrate polypyrimidine tracts, which are always located downstream of the branch point (Fig. 2). Exons in metazoan pre-mRNAs often contain enhancer-splicing elements (ESEs), usually purine-rich, which are recognized by specialized regulatory splicing factors. It is possible that GC-rich exon sequences in plant pre-mRNAs also act as recognition targets, although their involvement in any type of splicing regulation has not yet been established.

U12-type introns

A minor class of nuclear pre-mRNA introns, referred to as U12-type or AT–AC introns (because they frequently start with AT and terminate with AC) have recently been described. These introns contain different splice site consensus sequences, and are excised by an alternative U12-type spliceosome (Fig. 2). Their splicing also requires five snRNAs, of which only U5 is common to both spliceosome types, whereas U11, U12, U4atac and U6atac carry out the functions of U1, U2, U4 and U6 snRNAs, respectively. Other components of the splicing machinery appear to be shared by both spliceosomes. It is noteworthy that introns with GT–AG borders, which are spliced by the U12 spliceosome (Fig. 2), and introns with AT–AC borders, spliced by the classical U2 spliceosome also occur, at a frequency comparable to that of the U12-type with AT–AC termini. Hence, residues other than terminal dinucleotides determine which of the two spliceosomes will be used. U12 class introns represent ~0.1% of all introns. They are found in organisms ranging from higher plants to mammals, and their positions within equivalent genes are frequently phylogenetically conserved. The genomes of Saccharomyces cerevisiae and Caenorhabditis elegans do not contain U12-type introns. Because U12 introns clearly originated before the divergence of the plant and animal kingdoms, their absence in C. elegans is most easily explained by their conversion to U2-type introns or by intron loss, rather than by intron gain in plants and vertebrates. U12-type introns in plants are also enriched in U residues (Z.J. Lorković, D.A. Wieczorek Kirk, M.H.L. Lambermon and W. Filipowicz, unpublished).
The plant counterparts of both U2AF subunits have been cloned. Nevertheless, cloning of all spliceosomal snRNAs of the major class (U1, U2, U4, U5 and U6; Ref. 15) and some of the U12 class (U6atac, U12; Ref. 16) has indicated their similarity to metazoan counterparts in both their primary and secondary structure. All sequence elements necessary for either the assembly of snRNAs into RNP or for the snRNA–mRNA and snRNA–snRNA base-pairings taking place during the spliceosomal cycle are conserved in plant snRNAs. Likewise, based on sequence comparisons and limited in vitro reconstitution and immunological studies, most protein components of plant snRNPs appear to be conserved\(^\text{15,17,18}\). Interestingly, plant genomes generally encode many sequence variants of U1–U5 snRNAs. The expression of some of these might be developmentally regulated, but it is not known whether this has any biological significance\(^\text{15}\).

Inspection of the Arabidopsis Database, which to date contains \(~70\%\) of the genome sequence, indicates that like U-snRNPs, other protein factors participating in spliceosome assembly and splicing regulation are highly conserved. However, few of these plant proteins have been characterized experimentally. The best studied are a group of factors referred to as SR proteins. In metazoas, SR proteins play an important role in general and alternative splicing by promoting different types of intraplicosomal contacts and mediating the function of splicing enhancers. The SR proteins are composed of one or two N-terminally placed RBD-type RNA-binding domains (RBDs), interacting with specific sequences in the pre-mRNA, and a domain rich in Ser–Arg dipeptides, involved in protein–protein interactions\(^\text{11,19}\). Although most of the characterized plant SR proteins represent homologs of vertebrate proteins, some appear to be plant specific, containing domains without obvious counterparts in known human SR proteins\(^\text{22,25}\) (Table 1). Interestingly, SRZ-21, SRZ-22, SR33 and SR45 interact with the SR protein U2AF, a heterodimeric splicing factor composed of a large (U2AF 65) and a small (U2AF 35) subunit, binding to the 3’ ss-proximal polypyrimidine tract in mammalian introns and helps to position the U2 snRNP at the branch point located immediately upstream\(^\text{11}\). The plant counterparts of both U2AF subunits have been cloned.

snRNPs and protein factors
Biochemical characterization of the splicing machinery in plants has been hampered by the lack of systems that could recapitulate the splicing reaction in vitro. Nevertheless, cloning of all spliceosomal snRNAs of the major class (U1, U2, U4, U5 and U6; Ref. 15) and some of the U12 class (U6atac, U12; Ref. 16) has indicated their similarity to metazoan counterparts in both their primary and secondary structure. All sequence elements necessary for either the assembly of snRNAs into RNP or for the snRNA–mRNA and snRNA–snRNA base-pairings taking place during the spliceosomal cycle are conserved in plant snRNAs. Likewise, based on sequence comparisons and limited in vitro reconstitution and immunological studies, most protein components of plant snRNPs appear to be conserved\(^\text{15,17,18}\). Interestingly, plant genomes generally encode many sequence variants of U1–U5 snRNAs. The expression of some of these might be developmentally regulated, but it is not known whether this has any biological significance\(^\text{15}\).

Inspection of the Arabidopsis Database, which to date contains \(~70\%\) of the genome sequence, indicates that like U-snRNPs, other protein factors participating in spliceosome assembly and splicing regulation are highly conserved. However, few of these plant proteins have been characterized experimentally. The best studied are a group of factors referred to as SR proteins. In metazoas, SR proteins play an important role in general and alternative splicing by promoting different types of intraplicosomal contacts and mediating the function of splicing enhancers. The SR proteins are composed of one or two N-terminally placed RBD-type RNA-binding domains (RBDs), interacting with specific sequences in the pre-mRNA, and a domain rich in Ser–Arg dipeptides, involved in protein–protein interactions\(^\text{11,19}\). Although most of the characterized plant SR proteins represent homologs of vertebrate proteins, some appear to be plant specific, containing domains without obvious counterparts in known human SR proteins\(^\text{22,25}\) (Table 1). Interestingly, SRZ-21, SRZ-22, SR33 and SR45 interact with the SR protein U2AF, a heterodimeric splicing factor composed of a large (U2AF 65) and a small (U2AF 35) subunit, binding to the 3’ ss-proximal polypyrimidine tract in mammalian introns and helps to position the U2 snRNP at the branch point located immediately upstream\(^\text{11}\). The plant counterparts of both U2AF subunits have been cloned.

Recombinant forms of the two characterized large-subunit isoforms substitute functionally for the human counterpart in the HeLa cell splicing extract (Ref. 26, suggesting that the protein performs a similar function in mammals and plants. U2AF is assisted by other proteins that contact U2AF\(^\text{b}\) or U2AF\(^\text{c}\), contributing to the complex network of interactions necessary for the definition of introns and exons\(^\text{15}\). Sequences of some factors interacting with U2AF (UAP56, SF1) are present in the Arabidopsis Database.

Plant hnRNP proteins
Studies in mammals and insects have shown that nascent pre-mRNA transcripts emerging from the chromatin are bound by tens of different RNA-binding proteins, collectively referred to as hnRNP proteins. These proteins usually contain one or more RNA-binding domains of different types, and auxiliary domains, often with an unusual amino acid composition. The hnRNP proteins were originally thought to be responsible for packaging and for proper folding of processing substrates in the nucleus, but...
The Arabidopsis genome encodes two proteins with RBD domains that are similar to those found in the human hnRNP H/F proteins. In addition, plant hnRNP H regulates alternative splicing of some pre-mRNAs by binding to the splicing enhancers or silencers. Searches of the databases using other mammalian hnRNP proteins as queries, have not revealed obvious plant homologs. Some plant hnRNP-like proteins might have no counterparts in mammals, such as tobacco RZ-1 (Ref. 30) and the Nicotiana plumbaginifolia UBP1 protein, which interacts with the U-rich elements in introns (Ref. 31) both meet the criteria of hnRNP proteins (Fig. 4).

Role of U-rich sequences in intron recognition

Most research on plant pre-mRNA processing has focused on the role of intronic UA-rich sequences, a distinguishing feature of plant introns. It is now well documented that UA-rich sequences are essential for efficient splicing and for selection of the 5'ss and 3'ss (Refs 4–7). Experiments performed with UA-deficient introns have shown that short elements, such as UUUUAUAU or its multimers, activate splicing irrespective of their location (i.e. whether they are near the 5'ss, the 3'ss or if they are in the middle of the intron). Hence, these elements function differently from the metazoan polypyrimidine tracts, which generally act between the branch point and the 3'ss. Analysis of pre-mRNA processing in tobacco cells indicates that UA-rich elements also play a role in defining intron borders. The 3'ss and 5'ss preferentially selected for splicing are usually those present at the transition regions from UA- to GC-rich sequence. Mutational analysis of the UA-rich elements indicated that, both in splicing enhancers and in intron border definition, 10 to 14 U rather than A residues are important. This is consistent with the analyses of plant intron composition already discussed.

How, at the molecular level, do UA-rich sequences contribute to intron recognition during splicing in plants? One possible function could be to minimize the secondary structure of introns. Indeed, hairpins introduced into plant introns have a strong negative effect on splicing efficiency in dicot plant cells. However, their role as targets recognized by hnRNP-like proteins or related factors at early steps in intron recognition might be much more important. Proteins with specificity for oligouridylates that interact with intron sequences in vitro have been identified in nuclear extracts of *N. plumbaginifolia*. The intron-binding protein, named

### Fig. 2. Schematic representation of (a) signals important for splicing of the major, U2 class introns in yeast, vertebrates and plants and (b) the U12 class in metazoans and plants. The black boxes represent exonic splicing enhancers (ESEs); Plant ESEs remain to be characterized; the exonic GC-rich sequences might not necessarily represent ESEs. Numbers below plant splicing site consensus sequences represent the percentage occurrence of indicated bases at each position in dicot plants. Base pairings taking place between the 5'ss splice site (s) and the branch point signals in U12-dependent introns and small nuclear RNAs (snRNAs) are indicated. Arabidopsis (At) U6atac and U12 snRNAs have the same base-pairing potential as their human (Hs) counterparts (based on Refs 16 and 51). Note the extended consensus sequence at the 5'ss and branch point compared with the U2-dependent introns (Fig. 1). The distance between the branch point and 3'ss is shorter in U12 class introns (10–16 nt) than in U2 introns (20–40 nt). In both intron types, the branch point adenosine is marked with an asterisk. The polypyrimidine tract located between the branch point and 3'ss is prominent only in vertebrate introns. Programs for intron prediction that are available on the Internet include SplicePredictor (http://genetel1.oxford.ox.ac.uk/cgi-bin/sp.cgi), NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), MZEF (http://argon.cshl.org/genefinder/ARAB/ arab.htm) and GenScan (http://CCR-081.jmi.nctu.edu.tw/GENSCAN.html).

### Table 1. Plant intron composition already discussed.

<table>
<thead>
<tr>
<th>Vertebrates</th>
<th>Yeast</th>
<th>Plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'ss UACU</td>
<td>UACU</td>
<td>UACU</td>
</tr>
<tr>
<td>3'ss UG</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>BP GAAG</td>
<td>GA</td>
<td>GA</td>
</tr>
</tbody>
</table>

### Table 2. Essential RNA splicing factors at early steps in intron recognition.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>HnRNP</td>
<td>Essential for efficient splicing</td>
</tr>
<tr>
<td>SR</td>
<td>Essential for splicing at early steps</td>
</tr>
<tr>
<td>PTB</td>
<td>Essential for splicing at early steps</td>
</tr>
</tbody>
</table>

### Fig. 3. Like human PTB, the plant protein binds RNA in vitro, with specificity for poly(U) and poly(C). In metazoan, hnRNP A1/A2 and PTB proteins act as regulators of alternative pre-mRNA splicing. HnrNPs A1/A2 are antagonists of the SR protein ASPP2, stimulating use of the distal alternative 5'ss; PTB has been implicated in the selection of alternative 3'ss sites. Conservation of hnrNAP A/B and PTB proteins argues for an important role in pre-mRNA maturation in plants.
UBP1, contains three RBD-type RNA-binding domains and a glutamine-rich N-terminus. Consistent with a role in pre-mRNA processing, UBP1 is found in association with a nuclear poly(A) RNA in vivo, and its overexpression in protoplasts strongly enhances the splicing of otherwise inefficiently processed introns by an unknown mechanism. The binding of UBP1 to U-rich elements might delineate sequences to be excised as introns and maintain them in an open conformation suitable for other interactions. Alternatively, UBP1 might play a more active role, helping to attract other factors such as U-snRNPs to the splicing substrate. Interestingly, overexpression of UBP1 also increases the steady-state level of certain reporter RNAs, both intron containing and intronless, indicating that the protein targets multiple steps of pre-mRNA maturation in the nucleus.

The same protein does not necessarily recognize the U-rich elements present at different locations in the intron. Identification of plant homologs of the splicing factor U2AF (Ref. 26), which binds to the 3's proximal polyuridylate tract in metazoa, suggests that 3's proximal U-rich stretches in plant introns are recognized by this factor. It is also possible that the 3's proximal U-rich sequences are first recognized by the UBP1-like protein and then by the U2AF.

### Intron versus exon definition models

In mammals, exons rather than introns are the operational units that are defined at the early steps of pre-mRNA recognition by an unknown mechanism. The binding of UBP1 to U-rich elements might delineate sequences to be excised as introns and maintain them in an open conformation suitable for other interactions. Alternatively, UBP1 might play a more active role, helping to attract other factors such as U-snRNPs to the splicing substrate. Interestingly, overexpression of UBP1 also increases the steady-state level of certain reporter RNAs, both intron containing and intronless, indicating that the protein targets multiple steps of pre-mRNA maturation in the nucleus.

The same protein does not necessarily recognize the U-rich elements present at different locations in the intron. Identification of plant homologs of the splicing factor U2AF (Ref. 26), which binds to the 3's proximal polyuridylate tract in metazoa, suggests that 3's proximal U-rich stretches in plant introns are recognized by this factor. It is also possible that the 3's proximal U-rich sequences are first recognized by the UBP1-like protein and then by the U2AF.

### Intron versus exon definition models

In mammals, exons rather than introns are the operational units that are defined at the early steps of pre-mRNA recognition by an unknown mechanism. The binding of UBP1 to U-rich elements might delineate sequences to be excised as introns and maintain them in an open conformation suitable for other interactions. Alternatively, UBP1 might play a more active role, helping to attract other factors such as U-snRNPs to the splicing substrate. Interestingly, overexpression of UBP1 also increases the steady-state level of certain reporter RNAs, both intron containing and intronless, indicating that the protein targets multiple steps of pre-mRNA maturation in the nucleus.

The same protein does not necessarily recognize the U-rich elements present at different locations in the intron. Identification of plant homologs of the splicing factor U2AF (Ref. 26), which binds to the 3's proximal polyuridylate tract in metazoa, suggests that 3's proximal U-rich stretches in plant introns are recognized by this factor. It is also possible that the 3's proximal U-rich sequences are first recognized by the UBP1-like protein and then by the U2AF.

### Intron versus exon definition models

In mammals, exons rather than introns are the operational units that are defined at the early steps of pre-mRNA recognition by an unknown mechanism. The binding of UBP1 to U-rich elements might delineate sequences to be excised as introns and maintain them in an open conformation suitable for other interactions. Alternatively, UBP1 might play a more active role, helping to attract other factors such as U-snRNPs to the splicing substrate. Interestingly, overexpression of UBP1 also increases the steady-state level of certain reporter RNAs, both intron containing and intronless, indicating that the protein targets multiple steps of pre-mRNA maturation in the nucleus.

The same protein does not necessarily recognize the U-rich elements present at different locations in the intron. Identification of plant homologs of the splicing factor U2AF (Ref. 26), which binds to the 3's proximal polyuridylate tract in metazoa, suggests that 3's proximal U-rich stretches in plant introns are recognized by this factor. It is also possible that the 3's proximal U-rich sequences are first recognized by the UBP1-like protein and then by the U2AF.
3' and 5' splice sites do not need to rearrange before catalysis, in contrast with the exon definition scheme 36. However, recent analysis of several Arabidopsis splicing mutants have provided evidence of the operation of the exon definition scheme during splicing of at least some plant introns 37. Many mutations in the splice sites lead to exon skipping (i.e. splicing of the affected exon together with the two flanking introns) – the result predicted by the exon definition model. In addition, splicing of the inefficiently processed intron from the reporter RNA might be strongly stimulated by another intron positioned either up or downstream. This stimulation can also be achieved by the insertion of the functional 3' or 5' half of the intron at the up and downstream location, respectively, clearly demonstrating that interactions between the splicing factors are taking place across exons 38. It is likely that, as proposed for Drosophila 39, both intron and exon definition mechanisms are active in plants. Use of a particular pathway and the efficiency of intron processing would depend on the combined contribution of many factors. These include the quality of the splice sites and the branch point, the strength, position and density of the accessory cis-acting signals, such as intronic U-rich sequences; and probably, GC-rich sequences in exons. Processing would also depend on neighboring introns and exons, on their structure and length, and the properties of their own splicing signals 40. By analogy with mammalian splicing, interactions with the 5' cap-binding complex will probably be important for processing the 5' proximal intron, whereas communication with the polyadenylation machinery will affect the splicing of the most downstream intron in pre-mRNA. It should be emphasized that the relative contribution of individual splicing elements would vary between different introns. Consistent with this, some introns comprise poor-quality splice sites or do not conform at all to the UAGC intron and exon compositional bias 41,42.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Length (aa)</th>
<th>Plant</th>
<th>Localization or RNA binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>RZ-1</td>
<td>209</td>
<td>Nt</td>
<td>N+</td>
</tr>
<tr>
<td>UBP1</td>
<td>406</td>
<td>NprAf (3)</td>
<td>N+</td>
</tr>
<tr>
<td>hnRNP A/B</td>
<td>404-445</td>
<td>A (3)</td>
<td>N/ND</td>
</tr>
<tr>
<td>hnRNP A/B</td>
<td>431-460</td>
<td>A (3)</td>
<td>N/ND</td>
</tr>
<tr>
<td>hnRNP H/F</td>
<td>278</td>
<td>A (1)</td>
<td>ND/ND</td>
</tr>
<tr>
<td>hnRNP H/F</td>
<td>248</td>
<td>A (1)</td>
<td>ND/ND</td>
</tr>
<tr>
<td>hnRNP I (PTB)</td>
<td>-500</td>
<td>A (1)</td>
<td>ND/ND</td>
</tr>
</tbody>
</table>

Fig. 3. Schematic representation of plant hn ribonucleoprotein (hnRNP)-like proteins. Tobacco RZ-1 associates with the nucleoplasmic RNase-sensitive 60S RNP (Ref. 30). RZ-1 and UBP are cross-link to poly(A) in vivo 37. Six Arabidopsis proteins deduced from the genomic sequences (Accession nos Z97335, AB009052, AB025628, AC002332, AL078465 and AB015479) with strong similarity to metazoan hnRNP A/B are represented by two schemes, which reflect differences in their C-terminal auxiliary domain. Two proteins with similarity to metazoan hnRNP H/F proteins, deduced from the Arabidopsis genes (AB025629 and AB011474), are shown separately because they differ in the linker region between the RNA-binding domains (RBDs). The RNP1 and RNP2 motifs, [T/S]GEA[F/D]V[E/Q][F/L] and V[K/R/H][V/L/M]R, characteristic for metazoan hnRNP H/F, are also found in plant proteins. Arabidopsis proteins appear to contain only two RBDs instead of the three found in metazoan hnRNP H/F proteins 37. Like metazoan polypyrimidine tract-binding protein (PTB), the Arabidopsis homolog (AC005687) is composed of four RBDs. The number of different genes for each protein class in Arabidopsis, based on genomic sequencing, is indicated in parentheses. Abbreviations: aa, amino acid; N, Nicotiana tabacum; Nt, Nicotiana plumbaginifolia; At, Arabidopsis thaliana; N, demonstrated nuclear localization; +, demonstrated interaction with poly(A)+ RNA in vivo; ND, not determined.

Fig. 4. Intron and exon definition models for pre-mRNA recognition. The model is based on information derived from both metazoan and plant splicing. Interactions of U1 small nuclear ribonucleoprotein (snRNP, labelled U1) with the 5' splice site (ss), and U2AF (yellow) with the region between the branch point and the 3' ss are early events of splice-site recognition taking place within the E complex. Uridylate-specific RNA-binding hnRNP-like protein represented by green ovals, together with SR proteins (red ovals) bound to the putative exonic splicing enhancers, assist in the binding of U1 snRNP and U2AF to the pre-mRNA and in establishing cross-intron and cross-exon contacts between the splice sites. The branch point bridging protein (BRBP/SP1, blue) interacts with the branch point (A), helps to position U2AF to the polypyrimidine tract through cross-intron interactions 41. Because the gene encoding BRBP/SP1-like protein is present in Arabidopsis (AB020444), this interaction is likely to occur in the plant E complex as well. All components indicated, except exonic splicing enhancers (ESEs, black boxes), have been identified in plants. U1-70K depicted in gray.
As in mammals, SR proteins probably assist early interactions across exons and introns in plant pre-mRNAs. In mammals, SR proteins are thought to influence the use of upstream or downstream splice sites by contacting, either directly or indirectly, U2AF and U1 snRNP, respectively. SR proteins also contribute to interactions between U1 snRNP and the 3′ splice site across the intron11-13 (Fig. 4). UBPI and related proteins, which bind to U-rich elements dispersed throughout introns, probably represent factors that are important for establishing the cross-intron interactions in plant pre-mRNAs.

Alternative splicing

Primary transcripts of many genes are alternatively spliced, producing two protein isoforms that differ only at the C-terminus and have functional differences. Some alternative splicing events are constitutive, with similar ratios of variant mRNAs in different cells, whereas others are subject to tissue-specific or developmental regulation40. In plants, the number of known cases of alternative splicing is increasing although in most cases the biological significance is unknown41. A few examples of alternative splicing with documented biological relevance are shown in Figure 5. Other events and often quite complex alternative processing of transcripts originating from plant transposons and some DNA viruses have been discussed in reviews42-44.

Usage of alternative 5′ ss in the Rubisco activase pre-mRNA produces two protein isoforms that differ only at the C-terminus (Fig. 5). Although both isoforms can activate Rubisco, the larger one is repressed regulated. However, following premixing, reactivities in the larger isoform can effectively regulate the activity of both isoforms. It is interesting, that this type of regulation occurs in Arabidopsis and spinach but not in tobacco, which expresses only the shorter protein isoform, as do maize and Chlamydomonas43. Alternative 5′-splicing of the hydroxyprolyl-4

The presence of functional introns in transcription units can dramatically enhance mRNA accumulation, even by factors of hundreds, especially when the intron is placed close to the 5′ end of the gene. The extent of this intron-mediated enhancement (IME) depends on the origin of the intron, the flanking exonic sequences, the strength of the promoter used to drive RNA expression, and on the type and physiological state of the cells46-48. Strong IME has been observed mainly in monocots; in dicots it is usually not observed in monocots; in dicots it usually does not exceed two- to fivefold. However, a 30-fold IME of β-glucuronidase activity in Arabidopsis has been reported49. The molecular mechanism of intron-enhanced gene expression is not well understood.

Practical considerations and future prospects

Progress in sequencing Arabidopsis and other genomes has created a need to identify genes and their splicing patterns within large regions of uncharacterized DNA. Computer algorithms (Fig. 2 legend) can predict the location of introns from sequence factors, such as splice-site strength and coding potential. Plant splice-site prediction is much better characterized than in vitro approaches would aid

April 2000, Vol. 5, No. 4 165
between splicing in monocots and dicots and between different plant species. It would also aid the understanding of the principles behind alternative splicing, which is emerging as an important and widespread means to regulate gene expression in plants. In addition, because introns in many organisms other than plants are AU-rich, dissection of the basic splicing requirements would be interesting from an evolutionary point of view. Finally, the coupling of splicing with other processing reactions, and with transcription and RNA export, a subject of intensive studies in mammals and yeast (Ref. 50), is another important area that has yet to be addressed in plants.

References

Fig. 5. Schematic representation of some examples of biologically significant alternative splicing events in plants. The splicing patterns of (a) Rubisco activase and (b) HPR represent examples of alternative 5' splice site (ss) usage. (a) The alternative splicing of intron 6 results in the formation of a low level of longer mRNA, which encodes a shorter 43 kDa protein as a result of the premature stop codon (marked with an arrow). In pumpkin, alternative splicing of intron 12 in the HPR gene results in a protein, which contains a peroxisomal targeting signal at the C-terminus (underlined SKL). This alternative splicing event results in the production of mRNAs encoding two completely different proteins, ribosomal protein S14 and succinate dehydrogenase subunit B (SDHB). Both proteins contain a functional mitochondrial transit peptide (black box). Light gray boxes represent the exons of SDHB. (d) Alternative splicing of the tobacco mosaic virus (TMV) resistance gene causes the inclusion of a small exon present in intron 3 (gray box). The translated short protein N₁ confers complete resistance to TMV (Ref. 44). Arrows above the exons represent stop codons. Lengths of proteins are indicated. (e) Alternative processing of the FCA gene transcripts results in four different mRNAs: a, b, g and h. The only mRNA that encodes an apparently full-length FCA protein is g, which consists of two RBD-type RNA-binding domains and a WW domain involved in protein–protein interactions. The β-mRNA is produced by alternative cleavage and polyadenylation in intron 3. All introns of β mRNA are excised but it is alternatively spliced in intron 13. This results in a deletion of parts of exons 13 and 14 (black boxes) and termination of translation at a new in-frame stop codon at the start of exon 14 (Ref. 45).


