trends in plant science
Reviews

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.

Pre-mRNA splicing is one of the fundamental processes in constitutive and regulated gene expression in eukaryotes. During splicing, introns present in primary gene transcripts are removed and exons are ligated to produce translationally competent mRNAs. The basic mechanism of intron excision is similar in all eukaryotes. The reaction is mediated by the spliceosome, a large ribonucleoprotein (RNP) complex, which is assembled anew at each intron from small nuclear RNP particles (U-snRNPs) and numerous protein factors. Spliceosome assembly is a highly ordered and dynamic reaction, involving hydrolysis of several ATP molecules and many structural rearrangements (Fig. 1). Different classes of proteins, including RNA-binding proteins, RNA-dependent ATPases, RNA helicases and protein kinases, are responsible for the precision and regulation of this process. During splicing, exon and intron sequences have to be effectively distinguished and appropriate 5′ and 3′ splice sites (5′ ss and 3′ ss) have to be selected and juxtaposed prior to the catalytic step. The recognition of intron and exon borders requires a multitude of *cis*-acting elements, each with a relatively low information content, that are recognized by the *trans*-acting factors. More details on individual steps, structural rearrangements, and factors involved in the spliceosomal cycle can be found in recent reviews[1–3].

In this article, we review the properties of introns and the factors involved in intron recognition and spliceosome assembly in higher plants. We also discuss examples illustrating how gene expression in plants is regulated at the level of mRNA splicing. (Plant mRNA splicing is reviewed in Refs 4–7.)

Properties of plant introns

The intron and exon organization of higher plant genes is similar to that of vertebrates[4–7]. 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 are: 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, AG/GTAAGT and TGCAG/G, 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[4–10]. U-rich sequences, typically distributed throughout the entire length of plant introns, are required for efficient intron processing and splice-site selection[4–7]; 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[1,11]. 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[12] (Fig. 2).

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[3,13]. These introns contain different splice site and branch point 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, but 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[3,13,14]. 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[14]. 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).

1360 - 1385/00/$ – see front matter PII: S1360-1385(00)01595-8

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 RNPs 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[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[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 metazoa, SR proteins play an important role in general and alternative splicing by promoting different types of intraspliceosomal 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[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[20–25] (Table 1). Interestingly, SRZ-21, SRZ-22, SR33 and SR45 interact with the plant U1-70K protein, a structural component of U1 snRNP (Refs 22,25). At an early step in mammalian spliceosome assembly, within the E complex (Fig. 1), the U1 snRNP and the splicing factor U2 snRNP auxiliary factor (U2AF) are already positioned at the 5′ ss and the polypyrimidine tract, respectively. SRZ-22, SRZ-2, SR33 and SR45 might act as factors bridging the interactions between U1 snRNP and U2AF during splicing in plants. In the mammalian system, such bridging is essential for the progression of the spliceosomal cycle[11,19].

U2AF, a heterodimeric splicing factor composed of a large (U2AF^{65}) and a small (U2AF^{35}) subunit, binds to the 3′ ss-proximal polypyrimidine tract in mammalian introns and helps to position the U2 snRNP at the branch point located immediately upstream[1]. 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^{65} or U2AF^{35}, contributing to the complex network of interactions necessary for the definition of introns and exons[1,3]. Sequences of some factors interacting with U2AF (UAP56, SF1) are present in the *Arabidopsis* Database.

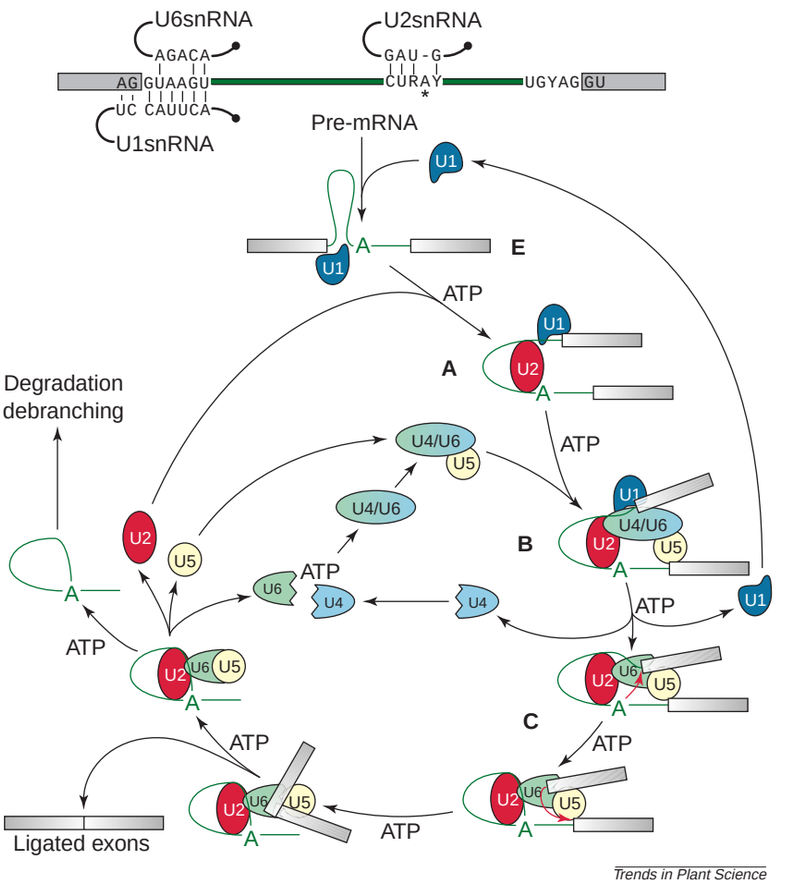

Fig. 1. The spliceosome cycle in mammals – this is probably also applicable to plants. Splicing of single intron-containing pre-mRNA is shown. Base pairings of the U1 and U6 small nuclear RNAs (snRNAs) with the 5′ splice sites (ss) and the U2 snRNA with the branch point are indicated based on conservation of these sequences in plant pre-mRNAs and snRNAs. Filled black circle at the ends of snRNAs represents a 5′ terminal cap structure. The involvement of five snRNPs and their interactions are indicated at distinct steps in spliceosome formation and catalysis. Except for complex E formation, all steps require ATP. Arrows indicate transitions between complexes and recycling of small nuclear ribonucleoprotein particles (snRNPs). After two transesterification reactions (red arrows), two ligated exons are released as mRNA. The intron lariat is released and subsequently debranched and degraded. Based on the mammalian cycle outlined in Ref. 3.

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

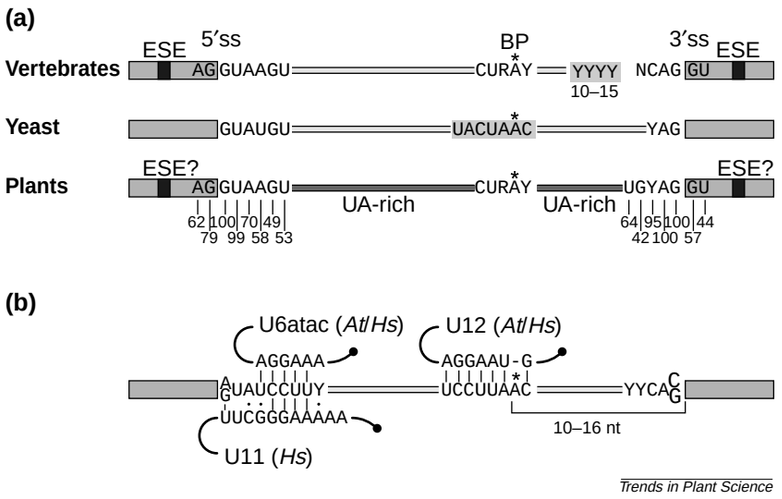

Fig. 2. Schematic representation of (a) signals important for splicing of the major, U2 class introns in yeast, vertebrates and plants and (b) the U12 class in metazoa and plants. The black boxes represent exonic splicing enhancers (ESEs). Plant ESEs remain to be characterized; the exonic GC-rich sequences might not necessarily represent ESEs. Numbers below plant splice site consensus sequences represent the percentage occurrence of indicated bases at each position in dicot plants. Base pairings taking place between the 5′ splice site (ss) 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://gremlin1.zool.iastate.edu/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.mit.edu/GENSCAN.html).

more recently many have been identified as factors involved in mRNA processing, transport, translation and stability. Several mammalian hnRNP proteins modulate alternative splicing patterns by binding, either alone or as parts of multiprotein complexes, to different elements in exons or introns. Interestingly, the same hnRNP protein can influence pre-mRNA maturation at several different levels. An example is hnRNP A1, which functions in alternative splicing and mRNA export[27,28].

Previous work suggested that nuclear extracts from plant cells contained few proteins interacting with pre-mRNA transcripts *in vitro*[29]. However, recent experiments and Database searches indicate that a complex set of hnRNP-like proteins is also expressed in plant cells (Fig. 3). For some of the listed examples, an association with poly(A)$^+$ RNA in the nucleus has been demonstrated directly. Six different 41–50 kDa proteins, with a striking similarity to the metazoan hnRNP A/B class proteins, are expressed in *Arabidopsis*. They comprise two RBDs, with 45–50% amino acid identity with metazoan counterparts, and an auxiliary C-terminal domain. As in most metazoan A/B proteins, the C-terminal domain in three plant proteins is Gly-rich; in three others, it is enriched in Asn, Gly and Ser. The plant genes are transcriptionally active in all major plant organs, and the encoded proteins are nuclear localized, consistent with a general role in pre-mRNA maturation (Z.J. Lorković, and W. Filipowicz, unpublished).

The *Arabidopsis* homolog of the hnRNP protein I, also known as the polypyrimidine tract-binding protein (PTB), has been identified (Fig. 3). Like human PTB, the plant protein binds RNA *in vitro*, with specificity for poly(U) and poly(C). In metazoa, hnRNP A1/A2 and PTB proteins act as regulators of alternative pre-mRNA splicing[27,28]. HnRNPs A1/A2 are antagonists of the SR protein ASF/SF2, stimulating use of the distal alternative 5′ ss; PTB has been implicated in the selection of alternative 3′ sites. Conservation of hnRNP A/B and PTB proteins argues for an important role in pre-mRNA maturation in plants.

The *Arabidopsis* genome encodes two proteins with RBD domains that are similar to those found in the human hnRNP H/F. H/F proteins interact with the nuclear cap-binding complex and are required for efficient RNA splicing *in vitro*. In addition, hnRNP H regulates alternative splicing of some pre-mRNAs by binding to the splicing enhancers or silencers[27,28]. 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 UUUUUAU 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[29]). 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[7,32,33]. Mutational analysis of the UA-rich elements indicated that, both in splicing enhancement[10,29] and in intron border definition[7,34], 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[35]. 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*[29]. The intron-binding protein, named

Table 1. Characterized plant spliceosomal proteins

Protein	Metazoan/yeast Homolog	Comment	Sequence motif(s)	Refs
snRNP components				
U1A	U1A/Mud1	Component of U1 snRNP	Two RBDs	17
U2B′′	U2B′′/Msl1	Component of U2 snRNP	Two RBDs	17
U1-70K	U1-70K/Snp1	Two alternatively spliced isoforms	RBD; RS, RD repeats	18
Splicing factors				
$U2AF^{65}$	$U2AF^{65}$/Mud2	Multiple forms expressed in plants; functionally replaces human $U2AF^{65}$; high affinity for poly(U) *in vitro*	Three RBDs; RS, RD repeats	26
$U2AF^{35}$	$U2AF^{35}$/none	Multiple forms expressed in plants	ψRBD; RS repeats; two putative Zn fingers	26
SR1(atSRp34)	SF2(ASF)/ none	Five alternatively spliced isoforms with differential expression during development	Two RBDs; RS, PS repeats	20,23
atRSp31	None	Recombinant protein complements human S100 extract	Two RBDs; RS, PS repeats	21
atRSp35	None		Two RBDs; RS, PS repeats	21
atRSp41	None		Two RBDs; RS, PS repeats	21
atSRp30	SF2(ASF)/ none	Similar to SR1/atSRp34; three alternatively spliced isoforms expressed differentially during development	Two RBDs; RS, PS repeats	23
SRZ-22 (atRSZp22) and SRZ-21	9G8,hSRp20/ none	Interact with U1-70K *in vitro* and in the yeast 2H system; complements human S100 and 9G8-depleted extracts, similar RNA-binding specificity to 9G8 and hSRp20	RBD; RS, PS repeats; Zn knuckle (CCHC); Gly-rich domain with five RGG repeats	22,24
SR33	SC35/none	Interacts with U1-70K, SR45 and with itself *in vitro*, and in the yeast 2H system	RBD; RS repeats	25
SR45	None	Interacts with U1-70K and with SR33 *in vitro*, and in the yeast 2H system	RBD; RS repeats at N- and C-terminus	25

Abbreviations: snRNP, small nuclear ribonucleoprotein particle; RBD, RNA binding domain containing RNP1 (octamer) and RNP2 (hexamer) submotifs; RS, PS, RD, RGG, domain rich in Arg–Ser, Pro–Ser, Arg–Asp and Arg–Gly–Gly repeats, respectively; U2AF, U2 snRNP auxiliary factor; 2H, yeast two-hybrid system.

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[31] 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[31].

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′ ss-proximal polypyrimidine tract in metazoa, suggests that 3′ ss-proximal U-rich stretches in plant introns are recognized by this factor. It is also possible that the 3′ ss-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 then introns are the operational units that are defined at the early steps of pre-mRNA recognition[36]. Such definition could occur by virtue of interactions across the exon, between factors bound at the 3′ ss and the downstream 5′ ss (Fig. 4). The fact that metazoan introns are frequently many kilobases long, whereas exons tend to be short (<300 nt), makes this mechanism appealing. Because plant introns are, by contrast, usually short and also harbor the recognition U-rich elements, their splicing is generally assumed to follow the intron-definition mechanism, with factors assembling at the 5′ and 3′ splice sites and interacting across the intron[5–7,9] (Fig. 4). An attractive aspect of this model is that complexes already positioned at the matching

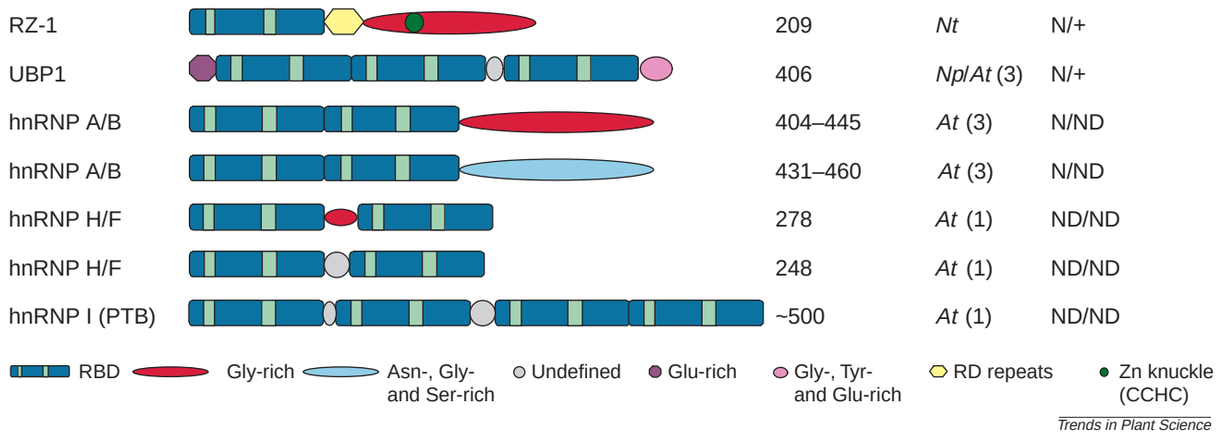

Fig. 3. Schematic representation of plant hn ribonucleoprotein (hnRNP)-like proteins. Tobacco RZ-1 associates with the nucleoplasmic RNase-sensitive 60S RNP (Ref. 30). RZ-1 and UBP1 cross-link to poly(A)$^+$ *in vivo*[31]. Six *Arabidopsis* proteins deduced from the genomic sequences (Accession nos Z97335, 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]RGL, 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[27]. 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 genome sequencing, is indicated in parentheses. Abbreviations: aa, amino acid; *Nt*, *Nicotiana tabacum*; *Np Nicotiana plumbaginifolia*; *At*, *Arabidopsis thaliana*; N, demonstrated nuclear localization; +, demonstrated interaction with poly(A)$^+$ RNA *in vivo*; ND, not determined.

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*[36], 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[4–7]. 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 UA/GC intron and exon compositional bias[5,6,39].

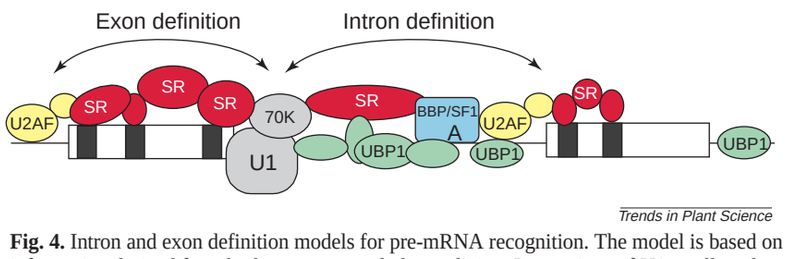

Fig. 4. Intron and exon definition models for pre-mRNA recognition. The model is based on information derived from both metazoan and plant splicing. Interactions of U1 small nuclear ribonucleoprotein (snRNP; labelled U1) with the 5′ splice site (ss), and U2AF (yellow) with the region between the branch point and the 3′ ss are early events of splice-site recognition taking place within the E complex. Uridylate-specific RNA-binding hnRNP-like proteins 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 (BBP/SF1; blue) that interacts with the branch point (A), helps to position U2AF[65] to the polypyrimidine tract through cross-intron interaction[1,3]. Because the gene encoding BBP/SF1-like protein is present in *Arabidopsis* (AB023044), 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 bind to ESEs and activate 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′ ss region across the intron[1,11,19] (Fig. 4). UBP1 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 different mRNA forms that encode proteins with 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 regulation[40]. In plants, the number of known cases of alternative splicing is increasing although in most cases the biological significance is unknown[5,6]. 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 reviews[4–6].

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, only the larger one is redox regulated. However, following premixing, redox changes 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 smaller protein isoform, as do maize and *Chlamydomonas*[41]. Alternative 5′-splicing of the hydroxypyruvate reductase (HPR) gene in pumpkin produces two different proteins with distinct subcellular localizations (Fig. 5). The HPR1 protein, but not HPR2, contains a C-terminal-targeting sequence for peroxisomes. Exposure to light greatly enhances the production of HPR2 mRNA, suggesting that alternative splicing might be light regulated in plants[42].

A rare example of a gene that is alternatively spliced to generate entirely different proteins is found in rice (Fig. 5). The two proteins, ribosomal protein S14 (RPS14) and succinate dehydrogenase subunit B (SDHB), are involved in mitochondrial protein synthesis and respiration, respectively. During evolution, the *rps14* gene probably moved from the mitochondria and became integrated into the intron of *sdhB*, which was already present in the nucleus; in this way, it has simultaneously acquired the capacity for nuclear expression and a mitochondrial targeting signal[43].

Recently, the biological relevance of alternative exon inclusion has been documented for the tobacco mosaic virus (TMV) resistance N gene (Fig. 5). The N gene produces two mRNAs, N_s and N_l, of which N_s is more abundant in non-infected plants whereas N_l becomes prevalent after infection. The N_l mRNA, resulting from inclusion of an alternative exon located in intron 3, encodes a shorter 652 amino-acid protein. This alternative splicing event is needed to confer complete resistance to TMV (Ref. 44).

Retention of an unspliced intron in a fraction of transcripts is common in plants, but probably reflects the low efficiency of normal splicing rather than a regulated process of any biological significance. However, the presence of an inefficiently spliced intron can potentially regulate the level of mRNA as a result of competition between splicing and polyadenylation[5,6]. A biologically relevant example is the processing of an RNA transcribed from the flowering control gene *FCA* in *Arabidopsis* (Fig. 5). Upon overexpression of the gene, flowering time is noticeably accelerated, an effect probably caused by a small increase in a fully spliced *FCA* mRNA. However, the level of the short transcript polyadenylated within intron 3, which is spliced inefficiently, increases >100-fold in transgenic plants, suggesting that this polyadenylation event controls the formation of FCA to keep flowering time constant[45].

Not much is known about the mechanisms underlying alternative splicing events in plants. In metazoa, SR proteins and other splicing factors and/or hnRNP proteins play an important role in regulating splice-site selection by binding to sequence elements (enhancers or silencers) present in exons and introns. Varying levels of specific SR proteins in different tissues might be responsible for the observed differences in mRNA splicing patterns[11,19,40]. The multitude of SR proteins identified in *Arabidopsis* suggests that these proteins might also act as splicing regulators in plants. Consistent with this expectation, plant proteins can modulate splice-site selection in the HeLa cell system *in vitro*[20,21,23,46]. More significantly, overexpression of atSRp30 in *Arabidopsis* modulates the alternative splicing of several endogenous mRNAs, including its own[23]. Proteins that are similar to the mammalian hnRNP proteins A/B and PTB possibly also act as regulators of splicing in plants. Physiological targets of plant SR proteins have not been identified. By analogy with metazoa, it is likely that GC-rich exons contain sequences acting as splicing enhancers. Indeed, an increase in exon GC content stimulates splicing of the neighboring intron in maize protoplasts[12].

Other regulatory roles of splicing

In metazoa, heat shock has a strong inhibitory effect on splicing, whereas plants frequently exposed to high temperatures in their natural environment might have evolved mechanisms to prevent the inhibition. However, other types of stress, such as exposure to cold, heavy metals or anaerobiosis, might affect the efficiency or patterns of splicing[4–6].

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 cells[4–7]. Strong IME has been observed mainly 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 reported[47]. 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 more accurate when the compositional contrast between exons and introns (Ref. 48) and a predicted branch point consensus are also considered (Ref. 49) (for the prediction of non-canonical U12-type introns, see Ref. 14). For the best results, it might be practical to use several programs and to compare the outputs.

Progress in understanding constitutive and regulated splicing requires the preparation of competent extracts of plant origin. With the wealth of *Arabidopsis* genome sequencing information now available, the *in vitro* approach would be particularly timely. Better characterization of *trans*- and *cis*-acting factors would aid the understanding of the basis for the reported differences

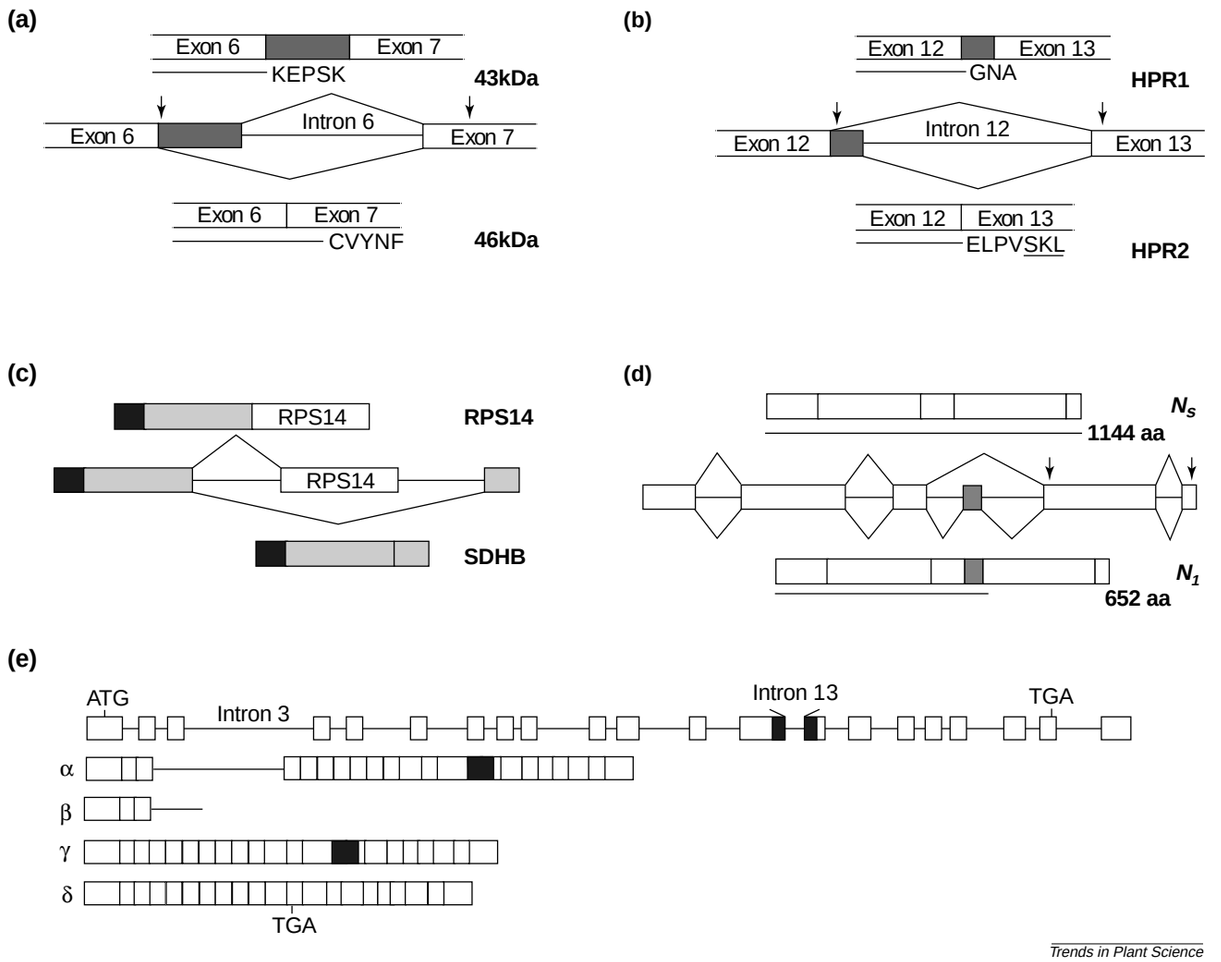

Fig. 5. Schematic representation of some examples of biologically significant alternative splicing events in plants. The splicing patterns of (a) Rubisco activase and (b) HPR represent examples of alternative 5′ splice site (ss) usage. (a) The alternative splicing of intron 6 results in the formation 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)[41]. 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)[42]. Dark gray boxes represent intronic sequences included as exons by alternative splicing. Arrows above the exons represent stop codons. The lines beneath the spliced mRNAs represent encoded proteins with few C-terminal amino acids indicated. (c) Alternative 3′ ss usage is exemplified by rice *sdhB*. This alternative splicing event results in the production of mRNAs encoding two completely different proteins, ribosomal protein S14 and succinate dehydrogenase subunit B (SDHB)[43]. 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 *N* gene causes the inclusion of a small exon present in intron 3 (gray box). The translated short protein N_l, 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: α, β, γ and δ. The only mRNA that encodes an apparently full-length FCA protein is γ, 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 δ 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).

between splicing in monocots and dicots and between different plant species[4–7]. 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 (Ref. 5), 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

1 Krämer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* 65, 367–409

2 Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell* 92, 315–326

3 Burge, C.B. *et al.* (1999) Splicing of precursors to mRNAs by the spliceosome. In *RNA World* (Gesteland, R.F. and Atkins, J.F., eds), pp. 525–560, Cold Spring Harbor Laboratory Press

4 Luehrsen, K.R. *et al.* (1994) Nuclear pre-mRNA processing in higher plants. *Prog. Nucleic Acid Res. Mol. Biol.* 47, 149–193

5 Simpson, G.G. and Filipowicz, W. (1996) Splicing of precursors to mRNA in higher plants: mechanisms, regulation and sub-nuclear organization of the spliceosomal machinery. *Plant Mol. Biol.* 32, 1–41
6 Brown, J.W.S. and Simpson, C.G. (1998) Splice site selection in plant pre-mRNA splicing. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 77–95
7 Schuler, M.A. (1998) Plant pre-mRNA splicing. In *A Look Beyond Transcription: Mechanisms Determining mRNA Stability and Translation in Plants* (Bailey-Serres, J. and Gallie, D.R., eds), pp. 1–19, American Society of Plant Physiologists
8 Goodall, G.J. and Filipowicz, W. (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58, 473–483
9 Brendel, V. *et al.* (1998) Intron recognition in plants. In *A Look Beyond Transcription: Mechanisms Determining mRNA Stability and Translation in Plants* (Bailey-Serres, J. and Gallie, D.R., eds), pp. 29–39, American Society of Plant Physiologists
10 Ko, C.H. *et al.* (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
11 Tacke, R. and Manley, J.L. (1999) Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* 11, 358–362
12 Carle-Urioste, J.C. *et al.* (1997) A combinatorial role for exon, intron and splice site sequences in splicing in maize. *Plant J.* 11, 1253–1263
13 Tarn, W-Y. and Steitz, J.A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22, 132–137
14 Burge, C.B. *et al.* (1998) Evolutionary fates and origins of U12-type introns. *Mol. Cell* 2, 773–785
15 Solymosy, F. and Pollak, T. (1993) Uridylate-rich small nuclear RNAs UsnRNAs, their genes and pseudogenes, and UsnRNPs in plants – structure and function – a comparative approach. *Crit. Rev. Plant Sci.* 12, 275–369
16 Shukla, G.C. and Padgett, R.A. (1999) Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* 5, 525–538
17 Simpson, G.G. *et al.* (1995) Molecular characterization of the spliceosomal proteins U1A and U2B′′ from higher plants. *EMBO J.* 14, 4540–4550
18 Golovkin, M. and Reddy, A.S.N. (1996) Structure and expression of a plant U1 snRNP 70K gene: alternative splicing of U1 snRNP 70K pre-mRNA produces two different transcripts. *Plant Cell* 8, 1421–1435
19 Valcárcel, J. and Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends Biochem Sci.* 21, 296–301
20 Lazar, G. *et al.* (1995) Identification of a plant serine–arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7672–7676
21 Lopato, S. *et al.* (1996) Characterization of a novel arginine/serine-rich splicing factor in *Arabidopsis*. *Plant Cell* 8, 2255–2264
22 Golovkin, M. and Reddy, A.S.N. (1998) The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell* 10, 1637–1647
23 Lopato, S. *et al.* (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes Dev.* 13, 987–1001
24 Lopato, S. *et al.* (1999) A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. *Plant Mol. Biol.* 39, 761–773
25 Golovkin, M. and Reddy, A.S.N. (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. *J. Biol. Chem.* 274, 36428–36438
26 Domon, C. *et al.* (1998) Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants. *J. Biol. Chem.* 273, 34603–34610
27 Weighardt, F. *et al.* (1996) The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *BioEssays* 18, 747–756
28 Krecic, A.M. and Swanson, M.S. (1999) hnRNP complexes: composition, structure and function. *Curr. Opin. Cell Biol.* 11, 363–371
29 Gniadkowski, M. *et al.* (1996) Characterization of intronic uridine-rich sequence elements acting as possible targets for nuclear proteins during pre-mRNA splicing *in Nicotiana plumbaginifolia*. *Nucleic Acids Res.* 24, 619–627
30 Hanano, S. *et al.* (1996) Isolation of a novel RNA-binding protein and its association with a large ribonucleoprotein particle in the nucleoplasm of tobacco cells. *Plant Mol. Biol.* 31, 57–68
31 Lambermon, M.H.L. *et al.* UBP1, a novel hnRNP protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO J.* (in press)
32 Lou, H. *et al.* (1993) 3′ splice site selection in dicot plant nuclei is position dependent. *Mol. Cell. Biol.* 13, 4485–4493
33 McCullough, A.J. *et al.* (1993) Factors affecting authentic 5′ splice site selection in plant nuclei. *Mol. Cell. Biol.* 13, 1323–1331
34 Merritt, H. *et al.* (1997) Internal AU-rich elements modulate activity of two competing 3′ splice sites in plant nuclei. *Plant J.* 12, 937–943
35 Liu, H-X. *et al.* (1995) Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5′ but not the 3′ splice site inhibit intron processing in *Nicotiana plumbaginifolia*. *EMBO J.* 14, 377–388
36 Berget, S.M. (1995) Exon definition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
37 Brown, J.W.S. (1996) *Arabidopsis* intron mutations and pre-mRNA splicing. *Plant J.* 10, 771–780
38 Simpson, C.G. *et al.* (1999) Interactions between introns via exon definition in plant pre-mRNA splicing. *Plant J.* 18, 193–302
39 Brendel, V. *et al.* (1998) Prediction of splice sites in plant pre-mRNA from sequence properties. *J. Mol. Biol.* 13, 85–104
40 Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32, 279–305
41 Zhang, N. and Portis, A.R. (1999) Mechanism of light regulation of Rubisco: a specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9438–9443
42 Mano, S. *et al.* (1999) Light regulates alternative splicing of hydroxypyruvate reductase in pumpkin. *Plant J.* 17, 309–320
43 Kubo, N. *et al.* (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternarive splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9207–9211
44 Dinesh-Kumar, S.P. and Baker, B.J. (2000) Alternatively spliced *N* resistance gene transcripts: Their possible role in tobacco mosaic virus resistance. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1908–1913
45 Macknight, R. *et al.* (1997) *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* 89, 737–745
46 Lopato, S. *et al.* (1996) Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. *Proc. Natl. Acad. Sci. U. S. A.* 93, 3074–3079
47 Rose, A.B. and Last, R.L. (1997) Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1. *Plant J.* 11, 455–464
48 Brendel, V. and Kleffe, J. (1998) Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in *Arabidopsis thaliana* genomic DNA. *Nucleic Acids Res.* 26, 4748–4757
49 Tolstrup, N. *et al.* (1997) A branch point consensus from *Arabidopsis* found by non-circular analysis allows for better prediction of acceptor sites. *Nucleic Acids Res.* 25, 3159–3163
50 Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* 11, 347–351
51 Dietrich, R.C. *et al.* (1997) Terminal intron dinucleotide sequences do not distinguish between U2 and U12-dependent introns. *Mol. Cell* 1, 151–160

Zdravko J. Lorković, Dominika A. Wieczorek Kirk, Mark H.L. Lambermon, and Witold Filipowicz* are at Friedrich Miescher Institute, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.
*Author for correspondence (tel +41 61 697 6993; fax +41 61 697 3976; e-mail filipowi@fmi.ch).