The removal of non-coding sequences, introns, from the mRNA precursors is an essential step in eukaryotic gene expression. U12-type introns are a minor subgroup of introns, distinct from the major or U2-type introns. U12-type introns are present in most eukaryotes but only account for less than 0.5% of all introns in any given genome. They are processed by a specific U12-dependent spliceosome, which is similar to, but distinct from, the major spliceosome. U12-type introns are spliced somewhat less efficiently than the major introns, and it is believed that this limits the expression of the genes containing such introns. Recent findings on the role of U12-dependent splicing in development and human disease have shown that it can also affect multiple cellular processes not directly related to the functions of the host genes of U12-type introns. At the same time, advances in understanding the regulation and phylogenetic distribution of the minor spliceosome are starting to shed light on how the U12-type introns and the minor spliceosome may have evolved.

© 2012 John Wiley & Sons, Ltd.

How to cite this article:
WIREs RNA 2012. doi: 10.1002/wrna.1141

INTRODUCTION

U12-type introns were initially described as a handful of unusual introns containing non-consensus AT–AC termini and a high degree of conservation at the 5′ splice site (5′ss). These characteristics set them apart from most other introns that typically had GT–AG termini and relatively variable sequences at the 5′ss. However, the true significance of this finding did not become evident until a provocative hypothesis by Hall and Padgett, who suggested that such introns may be recognized by the factors that are specific to this type of introns. This hypothesis set the stage for the discovery that the genomes of most eukaryotes actually harbor two different types of introns, termed U2 and U12 type, that are removed by two separate spliceosomes.

The significance of having two parallel intron types and two machineries for their removal was initially perplexing, especially since the U12-type introns are present at very low frequencies. The function and significance of the U12-dependent spliceosome are still far from being fully understood, but there are several lines of evidence indicating that U12-type splicing has essential functions. These are related to the conservation of U12-type introns in distantly related organisms and their presence in specific types of genes, as well as to the slow kinetics of their removal, which may regulate the expression of the genes containing them. Here, we review the current knowledge on U12-type introns and the function of the U12-dependent spliceosome. In particular, we discuss the recent findings on the regulation of the U12-dependent splicing, its function in development and disease, and the impact of these findings on understanding the evolution of the two spliceosomes.

U12-TYPE INTRON SEQUENCES: DEFINITION AND CHARACTERISTICS

Although the U12-type introns were first identified on the basis of terminal AT–AC dinucleotides (therefore initially referred to as AT–AC or atac introns) it was soon realized that these termini are not exclusively present in U12-type introns. Rather, the defining features of U12-type introns are the 5′ss and branch
point sequences (BPS), which are more conserved than in most other introns. GT–AG were also found to function as the terminal dinucleotides, and are actually the more common subtype. Thus, a nomenclature was adopted in which the original spliceosomal introns are referred to as U2-type introns and the novel introns as U12-type introns. The introns are spliced by the U2- or U12-dependent spliceosomes, and are also often referred to as the ‘major’ and ‘minor’ spliceosomes, respectively.

Naturally occurring U12-type introns mostly belong to either the AT–AC or GT–AG subtypes. A few natural examples of U12-type introns with other combinations of terminal residues have been reported, and non-canonical termini have also been shown to support U12-dependent splicing experimentally, although often with reduced efficiency. However, combinations of the main subtypes (i.e., AT–AG or GT–AC) appear to be disfavored. In addition to their distinct 5′ss and BPS, U12-type introns are also characterized by the lack of a distinct polypyrimidine tract (PPT) that is typically found upstream of U2-type 3′ splice sites. The distance from a U12-type BPS to the 3′ ss has been shown to be an important factor for the recognition of U12-type introns, and is significantly shorter than in U2-type introns, with an optimum distance of 11–13 nt.

DISTRIBUTION OF U12-TYPE INTRONS

Currently, U12-type introns have been identified in all major eukaryotic taxa, including plants, fungi, and animals, as well as a few deep-branching single-celled eukaryotes, but are nonetheless absent in many species, including such common model organisms as *Caenorhabditis elegans* and *Saccharomyces cerevisiae*. (Figure 2). The putative U12-type introns can readily be recognized computationally on the basis of the 5′ss and the BPS that are conserved among different organisms. An examination of the U12-type introns in the Splice Rack database (Ref 8; see Figure 1) revealed approximately 700–800 putative genes carrying U12-type introns in human and mouse, and ca 300 in *Arabidopsis thaliana*, while in *Drosophila melanogaster* only 19 U12-type introns have been described. However, these are estimates based on relatively stringent bioinformatic criteria, as there is little experimental evidence on the recognition of weak U12-type splice sites. Therefore, with less stringent criteria the number of putative U12-type introns could be significantly larger as suggested recently in *A. thaliana*.22

The genes containing U12-type introns are enriched in certain functional classes and pathways as described originally by Burge et al. They are mainly present in genes related to ‘information processing functions’, such as DNA replication and repair, transcription, RNA processing, and translation, but can also be found in genes related to cytoskeletal organization, vesicular transport, and voltage-gated ion channel activity. In contrast, U12-type introns are almost absent in genes related to basic energy metabolism and biosynthetic pathways. Typically only one U12-type intron is present in an individual gene. However, there are 50 genes containing two U12-type introns and a few cases with three U12-type introns. This is most common in genes belonging to the voltage-gated ion channel superfamily. Together, these observations have led to the suggestion that U12-type introns may have a role in regulating the expression of specific sets of genes.

SPLICEOSOME ASSEMBLY AND CATALYSIS

The Major Spliceosome

Introns are recognized and excised by a large molecular machine called the spliceosome that is composed of five small nuclear ribonucleoproteins (snRNPs) and additional non-snRNP protein components. Each of the five snRNPs contains one small nuclear RNA (snRNA) and a number of protein components. The U2-type snRNPs are U1, U2, U4, U5, and U6, the latter three of which form a U4/U6.U5 tri-snRNP. The spliceosome is formed by the sequential interactions of the snRNAs and spliceosomal proteins with the pre-mRNA substrate and with one another. In U2-type introns, the 5′ss is initially recognized by the U1 snRNP, while the BPS, PPT, and 3′ss are recognized by the protein factors SF1, U2AF65 and U2AF35, respectively, together forming the spliceosomal commitment (or E) complex. During the formation of the
Recognition of U12-Type Introns by the Minor Spliceosome

The U12-dependent spliceosome contains four specific snRNPs, U11, U12, U4atac, and U6atac, each of which contains a specific snRNA component that is equivalent to but distinct from its U2-type counterpart, i.e., U1, U2, U4, and U6, respectively. U5 snRNP is shared between the two spliceosomes. Although the sequences of the snRNAs with equivalent function are quite divergent in the two spliceosomes, they share a common overall secondary structure (Figure 4). U4atac, U6atac, and U5 associate into a tri-snRNP, similar to the major U4/U6.U5, and the protein composition of the major and minor tri-snRNPs appears to be very similar, if not identical. Moreover, it has also been shown that equivalent stem-loop structures of U4 and U4atac function as binding platforms for proteins that are required for the tri-snRNP formation. In contrast, while the major U1 and U2 snRNPs are distinct snRNPs, their counterparts are present in the nucleus as a U11/U12 di-snRNP. Interestingly, only seven protein components have thus far been reported to be specific to the minor spliceosomes (Table 1), and they are all located in the U11/U12 di-snRNP. U11/U12 also lacks all the U1-specific proteins and some U2-associated proteins (Table 1), making it the most divergent component of the minor spliceosome, in comparison to its counterparts in the major spliceosome.

The overall assembly pathways of the two spliceosomes are similar, and the main difference is the absence of a separate commitment complex in the minor spliceosome. Instead, the preformed U11/U12 di-snRNP binds the intron as a unit, and the 5′ss and 3′ss are recognized by U11/U12 and SF1, respectively, leading to the formation of the pre-spliceosome, or A complex, U2 snRNP replaces SF1 at the BPS. At later stages, the U4/U6.U5 tri-snRNP stably associates with the spliceosome (B complex). Rearrangements in RNA and protein interactions lead to the formation of the catalytically active spliceosome (B* complex) that catalyzes the first transesterification reaction, followed by the second transesterification in the C complex (Figure 3).
BPS are recognized in a cooperative manner within the A complex, although U11/5′ss basepairing still precedes the formation of stable U12/BPS basepairing. The initial basepairing interactions at the 5′ss are also different in the two spliceosomes: in contrast to the U1, the U11 snRNA does not basepair across the exon–intron boundary or even with the first three nucleotides of the intron. Instead, these nucleotides
The predicted secondary structures of the human spliceosomal snRNAs. The binding sites for Sm proteins are shaded in gray, and the sequences interacting with the 5' SS or BPS in cyan. Sequences involved in various U2/U6 or U12/U6atac interactions are indicated by green (helix I), purple (helix II), and yellow shading (helix III), similar to Figure 5. Nucleotide modifications are omitted. (Structures are based on data originally published in Ref 34 for U1, U2, and U5, Ref 18 for U11, Ref 35 for U12, and Ref 29 for U4, U6, U4atac, and U6atac). The locations and identities of the Taybi-Linder syndrome or microcephalic osteodysplastic primordial dwarfism type I (TALS/MOPD1) mutations in the U4atac snRNA are from Ref 36 and are indicated in red.
TABLE 1 | Proteins of the U1, U2, and U11/U12 snRNPs

<table>
<thead>
<tr>
<th>12S U1</th>
<th>17S U2</th>
<th>18S U11/U12</th>
<th>Functions (with Selected References)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sm proteins&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Sm proteins&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Sm proteins&lt;sup&gt;1&lt;/sup&gt;</td>
<td>snRNP core components&lt;sup&gt;26&lt;/sup&gt;</td>
</tr>
<tr>
<td>U1-A (SNRPA)</td>
<td>U1-C (SNRPC)</td>
<td>U1-70K (SNRNP70)</td>
<td>Structural; RNA-binding&lt;sup&gt;37&lt;/sup&gt;</td>
</tr>
<tr>
<td>U1-A (SNRPA1)</td>
<td>U2A&lt;sup&gt;′&lt;/sup&gt; (SNRPA1)</td>
<td>Structural; SR protein interactions&lt;sup&gt;37,38&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>U2B&lt;sup&gt;′&lt;/sup&gt; (SNRPB2)</td>
<td>SF3a complex&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Structural; RNA-binding</td>
<td></td>
</tr>
<tr>
<td>SF3b complex&lt;sup&gt;2&lt;/sup&gt;</td>
<td>SF3b complex&lt;sup&gt;2&lt;/sup&gt;</td>
<td>BPS binding&lt;sup&gt;39&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>20K (ZMATS)</td>
<td>25K (SNRNP25)</td>
<td>Unknown; homology to U1C&lt;sup&gt;32&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>31K (ZCRB1)</td>
<td>35K (SNRNP35)</td>
<td>SR protein interactions, homology to U1-70K&lt;sup&gt;33,41,42&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>48K (SNRNP48)</td>
<td>59K (PDCD7)</td>
<td>5′ss recognition&lt;sup&gt;43&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>65K (RNPC3)</td>
<td>Urp&lt;sup&gt;3&lt;/sup&gt; (ZRSR2)</td>
<td>Structural, binds 48K and 65K&lt;sup&gt;43,44&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>hPrp43&lt;sup&gt;4&lt;/sup&gt;</td>
<td>hPrp43 (DHX15)</td>
<td>3′ss recognition&lt;sup&gt;45&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Y Box-1&lt;sup&gt;3&lt;/sup&gt; (YBX1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Hugo names of proteins have been provided in parentheses.

1Sm proteins B/B’, D1, D2, D3, E, F, and G, 2Multi-subunit complexes.
3Also present in the major spliceosome, but not in U1 or U2 snRNPs.
4Almost stoichiometric presence in the 17S U2. Other proteins associated with U2 in substoichiometric amounts are omitted.<sup>46</sup>

are recognized specifically by the U11-48K protein, which likely also stabilizes the U11/5′ss helix.<sup>43,49</sup>

Surprisingly, even though U11-48K does not share significant sequence homology with the U1-specific protein U1-C, both appear to stabilize the binding of their respective snRNAs in a similar manner through their zinc finger domains.<sup>37,49</sup>

Interaction of the U12 snRNA with the BPS is similar to the U2/BPS basepairing in the major spliceosome, resulting in the exclusion of the branch point (BP) adenosine from the U12/BPS helix.<sup>50,51</sup> There seems to be flexibility in the choice of the BP adenosine, as either one of the two A residues present in the consensus BPS can be used as a BP, depending on the intron in question.<sup>52</sup> BPS recognition is likely stabilized by proteins of the SF3b complex, which is present in both spliceosomes (Table 1), and is known to bind to the BPS in the major spliceosome, with the protein factor p14 shielding the BP adenosine from premature activation.<sup>39,33,54</sup> U12-type introns do not have PPTs, and U2AF is not required for the recognition of U12-type introns.<sup>55</sup> However, a U11/U12 di-snRNP component related to U2AF<sup>35</sup>, Urp<sup>32</sup> is required for A complex formation and 3′ss recognition.<sup>45</sup> Urp is not specific to the minor spliceosome, but its function is different than in the major spliceosome, where it apparently displaces U2AF from the 3′ss after the first catalytic step.<sup>45</sup>

In the major spliceosome the U1 and U2 snRNPs recognize the 5′ss and BPS independently, and non-snRNP proteins are required for their association to each other.<sup>56</sup> In contrast, the U12-type 5′ss and BPS are linked through the internal components of the U11/U12 di-snRNP already at the earliest phase of intron recognition. The protein factors U11-59K and U11/U12-65K interact at the interface of the two snRNPs.<sup>44</sup> U11-59K further interacts with the U11-48K protein,<sup>43</sup> while U11/U12-65K directly binds to the 3′ terminal stem loop of U12 snRNA.<sup>44</sup> Owing to the compact structure of the di-snRNP,<sup>57</sup> further protein and/or RNA interactions are likely to contribute to the association of the two snRNPs. Furthermore, the binding sites for the 5′ss and BPS must be close to each other within the U11/U12 structure, as the very 5′ end of U12 snRNA can be cross-linked to the pre-mRNA 2 nt upstream of the 5′ss already in the A complex, suggesting that the BPS and 5′ss must be within 50 Å of one another.<sup>58</sup> Thus, the 5′ end of
U12 is already close to the position required for the formation of the catalytic core of the spliceosome.

Assembly of the Catalytic Core is Similar in the Two Spliceosomes

The formation of the catalytically active spliceosome is thought to follow a pathway similar to that of the major spliceosome. After initial recognition of the splice sites by the U11/U12 di-snRNP, the U4atac/U6atac.U5 tri-snRNP associates with pre-spliceosome to form complex B. A terminal stem-loop structure in the U6atac snRNA (possibly together with U4atac snRNA sequences) contains a signal that directs the U4atac/U6atac.U5 tri-snRNP to the minor and not to the major pre-spliceosome. A large number of structural rearrangements convert the pre-catalytic spliceosome (complex B) to the catalytic configuration (complexes B* and C) in a manner similar to the major spliceosome. U6atac snRNA replaces U11 at the 5′ss, U4atac/U6atac structure is unwound, and U12 and U6atac basepair with each other to form the ‘catalytic core’ structure in which the reactive A residue at the BP and the 5′ss are juxtaposed for the first step of the catalysis.

Although the formation of the catalytic core appears to be somewhat more flexible in the U12-dependent spliceosome, the structure and function of the core are likely to be highly similar. Indeed, both spliceosomes employ the same two-step transesterification mechanism for intron removal and U12-dependent splicing is even supported by the modified U6atac snRNAs in which the functional domain has been replaced by that of U6 snRNA. The U12/U6atac interactions in the catalytic core of the U12-dependent spliceosome resemble those in the major spliceosome, although the helix II structure present in the U2/U6 complex cannot be formed in the minor spliceosome. Nevertheless, the similarity of snRNA structural domains suggests that, like the major spliceosome, the minor spliceosome is likely to use RNA-based catalysis.

Removal of U12-Type Introns Is Slow

Both intron types are spliced co-transcriptionally within the nucleus and might be expected to display similar kinetics. However, the removal of U12-type introns appears to be significantly slower. Early in vitro splicing experiments documented a splicing rate for U12-type introns that was three- to fivefold slower than that of U2-type introns. Similar observations were also made in in vivo experiments, where approximately twofold higher levels of unspliced U12-type introns were detected in the steady-state transcript pools isolated from insect and mammalian cells. This is consistent with the observation that co-transcriptional splicing of U12-type introns is at least twofold slower than that of U2-type introns. Thus far, the reason for the lower efficiency remains unknown.

![Figure 5](image-url)
spliceosome snRNPs are ca 100-fold less abundant than the major snRNPs, and this could underlie the observed kinetic differences. However, the observation that a further 10-fold reduction in the levels of U4atac snRNA has no apparent effect on the efficiency of endogenous U12-dependent splicing argues against this simplistic explanation. The slower rate could also be related to the kinetic effects caused by the less flexible recognition phase of U12-type introns or by the inability to form some of the structures present in the catalytic core of the major spliceosome. Regardless of the underlying mechanism, Patel et al. have suggested that the slower rate of splicing could constitute a rate-limiting mechanism for the expression of genes containing U12-type introns. In this model, the transcripts containing unspliced U12-type introns would get trapped in the nucleus where they could be targeted by nuclear surveillance mechanisms.

EXON DEFINITION INTERACTIONS AND ALTERNATIVE SPlicing

Splicing enhancer and silencers sequences are important for defining splice sites both in constitutive and alternative splicing, and typically function by recruiting regulatory splicing factors, such as SR and hnRNP proteins. U12-type introns are not an exception in this regard, as U12-dependent splicing also responds to SR and hnRNP proteins that bind to splicing regulatory elements near the U12-type splice sites. A prototypical case of SR protein-mediated enhancement in the major spliceosome is the interaction of SRSF1 with the U1-70K protein, which enhances 5′ss recognition. The U11/U12 di-snRNP protein U11-35K is the putative paralog of U1-70K, and may have a similar function in stabilizing U11/5′ss binding. This notion is indirectly supported by the findings that U11-35K interacts with SR proteins in plants, including SRSF1 homologs, and that U12-type A complex formation is enhanced by SR proteins. As in the major spliceosome, SR proteins likely enhance intron recognition in multiple ways and through several spliceosomal factors, and have also been shown to directly contact U12-type 5′ss and BPS.

A particular problem related to the processing of nascent transcripts is that introns (especially those of vertebrates) can be extremely long, such that it can take an hour for an entire intron to be transcribed. Therefore, initial splice site definition typically takes place over the considerably shorter exons, with 3′ss-recognizing factors in the upstream intron interacting with the 5′ss-recognizing factors in the downstream intron (Figure 6(a)). Again, U12-dependent splicing has been shown to be enhanced by the presence of flanking U2-type introns, an effect which is likely mediated by the SR proteins or other auxiliary factors. The most detailed evidence for exon definition interactions between components of the two spliceosomes comes from the alternative splicing (AS) of the transcripts coding for the minor spliceosome components U11-48K and U11/U12-65K. In this system, an alternative U2-type 3′ss is activated by the U11/U12 di-snRNP bound at an atypical downstream splicing enhancer element consisting of a tandem repeat of U12-type 5′ss, designated as U11 snRNPP-binding splicing enhancer (USSE; Figure 6(b)). The alternatively spliced transcripts are degraded by the nonsense-mediated decay (NMD) pathway or other RNA degradation machineries, and this system thus functions as a regulatory feedback mechanism for minor spliceosome components.

Given that AS is extremely prevalent in humans, and the extensive interactions of the minor spliceosomes with other splicing factors, it is somewhat surprising that AS events involving U12-type introns appear to be almost entirely absent. This may be because of the more rigid sequence and distance constraints on the U12-type 5′ss, BPS, and
3’ss. AS events involving the use of an alternative 3’ss have been observed, but it is not entirely clear whether they are the result of true regulated AS or of splicing errors, especially since the selection of the U12-type 3’ss has been shown to be prone to errors.13,16,82,83 Exon skipping, which is the most common type of AS observed with major introns, has been reported only in a few instances involving U12-type introns, and the functional significance of these AS events is not known.14,16 The absence of exon skipping is likely to be due to the general incompatibility of U12 and U2-type splice sites and the scarcity of genes containing more than one U12-type intron. Alternative usage of mutually exclusive U12-type and U2-type splice sites has been observed in at least three cases. In vertebrates, the members of the JNK gene family contain a hybrid intron with a U12-type 5’ss and U2-type BPS and 3’ss. These can be spliced in a mutually exclusive fashion to either the 3’ss of the downstream U12-type intron or to the 3’ss of the upstream U2-type intron, respectively,82 and at least in mice these isoforms show tissue specificity, with the former isoform preferentially expressed in neurons. The D. melanogaster genes prospero and dUrp (which encodes the Urp protein present in both spliceosomes) contain overlapping U12 and U2-type introns or ‘twintrons’.2,14,84 The U12-dependent dUrp mRNA isoform is apparently destroyed by NMD,14 while prospero alternative splicing results in distinct developmentally regulated protein isoforms with amino-acid differences in the homeodomain region.85,86

PHYSIOLOGICAL SIGNIFICANCE OF THE MINOR SPLICEOSOME

Minor Splicing Is Essential for Development

Knockdown of protein components specific to the minor spliceosome leads to reduced proliferation of cultured cells,32,43 and thus minor splicing is an essential process. Proper U12-dependent splicing is also required for development, as shown in humans87,88 and in D. melanogaster,84 despite the fact that the latter only contains a handful of U12-type introns.20,21 Blocking minor splicing in zebrafish embryos resulted in developmental defects at a distinct stage, in comparison to blocking the major spliceosome.73 Knockdown of the U11-31K protein in A. thaliana also led to abnormal growth of the plant, but only after the bolting stage.40 These findings indicate that minor splicing has specific functions in regulating developmental processes. These functions may be mediated in part by alternative splicing, as with the prospero transcripts in D. melanogaster. However, it is likely that another, even more significant factor affecting developmental processes is the slower kinetics of minor splicing, which can limit not only the expression of the genes containing them68,70,71 but also downstream pathways. Indeed, defects in minor splicing have been shown to perturb the expression of multiple metabolic genes (which themselves do not contain U12-type introns) in D. melanogaster, likely underlying the developmental arrest.84,89 The developmental defect following the U11-31K knockdown in A. thaliana was also associated with downregulation of a specific set of genes related to gibberellic acid metabolism.40 Thus, U12-dependent splicing is necessary for multiple processes required for the viability and development of multicellular organisms, and can also affect specific cellular pathways.

Minor Splicing and Human Disease

From a medical point of view, it would be interesting to know to what extent the activity of the minor spliceosome contributes to regulating the expression of genes in different human tissues. This question remains mostly unanswered. Some protein components specific to the minor spliceosome (31K and 65K)
have been observed to be differentially expressed in different tissues,\textsuperscript{90,91} and U12-type intron-containing genes in bone marrow CD34 positive cells and B lymphoblasts,\textsuperscript{24} but the correlation between minor splicing and gene expression in different tissues has not been analyzed systematically. To date, only a few diseases have been linked to defects in minor splicing, some of them displaying tissue-specific symptoms, while others have a wide range of system-wide defects. Mutations in U12-type 5'ss in the LKB1 and SEDL genes cause Peutz-Jeghers syndrome and spondyloepiphyseal dysplasia tarda, respectively,\textsuperscript{83,92,93} and these diseases likely arise from the inactivity or insufficiency of the respective gene products.

A more complex class of diseases are those arising from defects in minor spliceosome components of which the TALS/MOPD1 (Taybi-Linder syndrome or microcephalic osteodysplastic primordial dwarfism type I) is a recent example. This severe developmental disorder is caused by recessive mutations in the U4atac gene (Refs 87 and 88; see Figure 4). Unlike their counterparts in the major spliceosome, each minor snRNA is expressed from a single locus in the genome,\textsuperscript{88,94,95} and the mutations in such loci can therefore potentially disable the function of the snRNA in question. In the case of TALS/MOPD1, the mutations are mostly in the 5' stem loop of U4atac (shown in Figure 4) and are predicted to prevent the binding of tri-snRNP-specific 15.5K and 61K proteins which could lead to a defect in the formation of U4atac/U6atac.U5 tri-snRNP,\textsuperscript{36,96} but experimental evidence for this is still lacking. This presumed defect in snRNP assembly decreases the activity of the U12-dependent spliceosome almost 10-fold,\textsuperscript{98} but still allows correct splicing of U12-type introns in most genes, albeit at significantly reduced levels. This resembles a situation with D. melanogaster carrying mutations in U6atac snRNA.\textsuperscript{70} An open question is whether the developmental defects observed in this disorder are due to a large-scale reduction in the expression of U12-type intron-containing genes or just a few affected genes in each tissue as has been suggested for D. melanogaster.\textsuperscript{89} A comparable case is spinal muscular atrophy (SMA), which arises from the mutation of the gene coding for the general snRNP assembly factor SMN, and is characterized by multiple defects in mRNA processing, including widespread changes in alternative splicing,\textsuperscript{97,98} but initially only affects spinal cord neurons. SMA has been reported to involve a decrease in the levels of the minor tri-snRNP\textsuperscript{99} and possibly U11/U12 di-snRNP levels as well,\textsuperscript{98,100} together with defective splicing of a few, but not all, U12-type introns.\textsuperscript{99} While the direct link of these findings on the pathogenicity of SMA is not entirely clear, it is possible that the defect in the splicing of some U12-type introns in neurons either causes or contributes to the disease phenotype.

**EVOLUTION OF THE MINOR SPLECSOME**

It is clear that the U12-dependent spliceosome is crucial for the viability and development of many multicellular organisms. However, intriguing questions about the minor spliceosome remain: Where did it originate? Why is it not present in all eukaryotes, and do its function and importance vary in different evolutionary lineages? We do not currently have clear answers to these questions, but recent advances in large-scale sequencing and bioinformatics are paving the way toward deeper understanding of the significance of the minor spliceosome.

**Origins and Conservation of the Minor Spliceosome**

The two spliceosomes have been proposed to originate from group II self-splicing introns, but it is not clear whether they arose simultaneously or sequentially, and whether this took place in one ancestral eukaryote or separate lineages that later fused.\textsuperscript{4,101–103} However, based on the distribution of introns and spliceosome components, it is evident that both U2- and U12-type introns and spliceosomes were present in the last common eukaryotic ancestor (Refs 18, 19, and 23; Figure 2). U12-type introns show remarkable conservation over long evolutionary distances: the rate of intron loss or conversion has been very low in all studied vertebrates, and while D. melanogaster has only 19 U12-type introns, almost all of them have orthologous U12-type introns in human.\textsuperscript{14} Most remarkably, 20 U12-type introns are conserved in orthologous positions between human and A. thaliana.\textsuperscript{16,104} Also, although the sequences of U12-type snRNAs are more diverged in distantly related organisms than their U2-type counterparts, their structures are highly conserved.\textsuperscript{18,20,41,67}

**Repeated Loss of U12-Type Introns**

Despite the high overall conservation of U12-type introns, they are entirely absent from many eukaryotes whose relatives nonetheless have them, suggesting repeated loss of U12-dependent splicing during evolution. The loss seems to have been most prevalent among diverse eukaryotic microbes, but some animals, including some but not all nematodes, have also lost...
their U12-type introns (Refs 18 and 19; Figure 2). The loss of U12-type introns has also been extensive in Diptera, which typically only have 15–20 U12-type introns in total in their genomes, several fold less than in most other insects.14,21 The precise reasons behind these losses remain enigmatic. However, it has been suggested that introns confer a slight deleterious effect for the fitness of the organism,101 and could thus be removed more efficiently in organisms with large effective population sizes, such as microbes and small invertebrates.14 Presumably, U12-type introns may be even slightly more deleterious, due to their slower excision and lower accuracy. In support of this, the loss of U12-type introns generally correlates with reduction in overall intron frequency, but U12-type introns still tend to be lost to a larger extent.14,21

In most cases, the mechanism of U12-type intron loss involves a simple deletion of the intronic sequence, apparently through the homologous recombination of reverse transcribed mRNAs.14 Another pathway is the step-wise conversion of U12-type introns into U2-type introns. Natural examples as well as experimental approaches have also shown that simple point mutations in the 5′ss of a U12-type intron, especially in the GT–AG subtype, suffice to turn it into a U2-type 5′ss,4,5,14 as U2-type introns have much more degenerate splicing signals. Subtype switches from a U12-type AT–AC intron to U12-type GT–AG are also observed.4,8,14 However, such changes appear to be rare, possibly due to the poor splicing efficiency of the intermediate forms (e.g., AT–AG).11,17,21,105 Sometimes intron conversion can also occur by the activation of a cryptic U2-type 5′ss near the original U12-type 5′ss.16,19 Given that the chance of creating the highly stringent U12-type 5′ss from a U2-type 5′ss is low, conversion in the reverse direction is a highly unlikely event. In fact, only one instance of a novel U12-type intron has been observed (in the dUrp twintron).14 Thus, without natural selection acting in favor of maintaining U12-type introns, they are likely to be converted into U2-type introns during evolution.

**Conservation of Regulatory Features**

Many splicing factors are regulated through a feedback mechanism, in which they bind to regulatory elements within their own pre-mRNA, resulting in alternatively spliced, unstable mRNA isoforms. Interestingly, all canonical SR and hnRNP proteins, as well as a number of core spliceosomal factors, are regulated in this manner, and in many cases the regulatory elements are also evolutionarily conserved.24,106–110 While most of the splicing factors thus regulated are common to both spliceosomes, certain spliceosome-specific proteins have also been shown to be alternatively spliced to produce unstable mRNA isoforms. These include, e.g., U1-70K and U11/U12-25K, for which the AS events are also evolutionarily conserved between human and mouse.108 The conservation of the AS events suggests that they have an important function in regulating the levels of these proteins, and may consequently affect the activity of the specific spliceosome in question. The feedback mechanism regulating the levels of minor spliceosome-specific proteins U11-48K and U11/U12-65K through the USSE element (Figure 6(b)) is conserved over extremely long evolutionary distances, in plants as well as in animals.81 Interestingly, the USSE is most conserved in vertebrates,31 which also show the most extensive conservation of U12-type introns,8,14,105 and is absent in Diptera, where both the overall numbers and interspecies conservation of U12-type introns are low.21 It is thus possible that this mode of regulation is important for the rate-limiting function of the minor spliceosome, without which the presence of U12-type introns would be mainly deleterious. Interestingly, the emergence of an AS-NMD-targeting U12-type intron in the Drosophila dUrp14 suggests that it, too, may function as a feedback mechanism for the minor spliceosome. Thus, it is possible that mechanisms limiting the activity of the minor spliceosome are essential for its function in regulating the expression of U12-type intron-containing genes.

**CONCLUSION**

Historically, insight into the function and significance of the U12-dependent spliceosome has come primarily from the biochemical characterization of splicing factors and spliceosome assembly, and has recently been complemented by the increasing number of bioinformatic analyses. However, our understanding of the significance of minor splicing is far from complete in both regards. The function of many of the U12-specific components is still unknown, and more components might still remain undetected. Importantly, although there are many independent lines of evidence indicating that U12-type introns are important for the regulation of gene expression, definitive large-scale evidence on the connection between the activity of the minor spliceosome and gene regulation is still lacking. Are there more regulatory elements that respond to U12-type factors? In what other ways is the minor spliceosome regulated?
and how does this correlate with the expression of U12-type intron-containing genes in different tissues? Could there be U12-type introns with more degenerate splice sites, and could they be involved in more extensive alternative splicing? An exciting way of studying these questions is now offered by the rapidly developing high-throughput sequencing and cross-linking methods, and will hopefully deepen our understanding on the significance of minor splicing in the coming years.

ACKNOWLEDGMENTS
This work was supported by the Academy of Finland and Sigrid Jusélius Foundation (MJF) and by the Helsinki Graduate Program in Biotechnology and Molecular Biology (EHN).

REFERENCES


