Acta Biologica Hungarica 54 (3–4), pp. 219–231 (2003)

HOW TO FIND THE REAL ONE
(AT THE LEVEL OF PRE-mRNA SPLICING)

Review

T. RAUCH and IBOLYA KISS*

Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences,
P.O. Box 521, H-6701 Szeged, Hungary

(Received: October 21, 2002; accepted: December 1, 2002)

The mature mRNA always carries nucleotide sequences that faithfully mirror the protein product according to the rules of the genetic code. However, in the chromosome, the nucleotide sequence that represents a certain protein is interrupted by additional sequences. Therefore, most eukaryotic genes are longer than their final mRNA products. The human genome project revealed that only a tiny portion of sequences serves as protein-coding region and almost one quarter of the genome is occupied by non-coding intervening sequences. The elimination of these non-coding regions from the precursor RNA in a process termed splicing must be extremely precise, because even a single nucleotide mistake may cause a fatal error. At present, two types of intervening sequences have been identified in protein-coding genes. One of them, the U2-dependent or major-class is prevalent and represents 99% of known sequences. The other one, the so-called U12-dependent or minor-class of introns, occurs in much lesser amounts in the genome. The basic problem of nuclear splicing concerns i/ the molecular mechanisms, which ensure that the coding regions are correctly recognized and spliced together; ii/ the principles and mechanisms that guarantee the high fidelity of the splicing system; iii/ the differences in the excision mechanisms of the two classes of introns. We are going to present models explaining how intervening sequences are accurately removed and the coding regions correctly juxtaposed. The two splicing mechanisms will also be compared.

Keywords: Pre-mRNA splicing – spliceosome – U2-dependent intron – U12-dependent intron – exon definition model

INTRODUCTION

The protein-coding regions transcribed by the RNA polymerase II (RNA pol II) are interrupted by introns that are removed from the pre-mRNA by a multi-step process termed splicing. In this process the coding sequences are spliced to generate the mature mRNA which is subsequently transported to the cytoplasm and translated into protein.

One of the most astonishing results of human genome analysis is that only ~1% of the total genome consists of protein-coding sequences (exons), while intervening

* Corresponding author; e-mail: kiss@nucleus.szbk.u-szeged.hu

sequences (introns) occupy ~24% of the genome, and the remaining chromosomal portions contain repetitive and other intergenic sequences [53]. On the basis of statistical data, an average human protein-coding gene is ~27 kilobases long and has 9 exons. The 5' and 3' terminal exons at each end of the pre-mRNA are longer than the internal ones, their average lengths are ~500 nucleotides (nt). The theoretical internal exons are relatively short (~145 nt), whereas the introns come up to several thousand nucleotides in length (~3400 nt). In practice, intron size can extend up to more than 200,000 nt. The correct removal of intronic sequences from pre-mRNA is a prerequisite for the generation of functional proteins. Regions that are essential for intron removal, referred to as splice sites, are located at the exon/intron boundaries. These are short and strictly, but not absolutely conserved *cis* elements. As many copies of similar potential sequences are carried by the same mRNA precursor, a very precise mechanism must be responsible for splice site selection. This is the so-called splice site choice enigma, which involves two additional problems: recognition of the real splice sites and ligation of the proper partners across the intron. To meet these requirements, a highly organized ribonucleoprotein complex (spliceosome) is formed on the *cis* elements of the RNA template for the elimination of the intronic region.

THE SPLICING REACTION OF EUKARYOTIC GENES

1. Players of the splicing machinery

1.1. Cis RNA elements

Right after the discovery of pre-mRNA splicing, common sequence features of the splice sites have been recognized [3]. Thus consensus *cis* elements were found at the two ends of introns (named 5' [5`ss] and 3' [3`ss] splice site) and 18–38 nt upstream of the 3' end (branch site) (Fig. 1A). Referring to the invariant GT and AG dinucleotides at the ends of introns (corresponding to GU and AG in the RNA), the intervening sequences of this type were called GT-AG introns representing the major class of presently known introns. In addition to this, other *cis* elements serve as recognition sites for protein factors (e.g. hnRNPs, SR proteins) participating in the splicing reaction, RNA transport or protection [13].

1.2. Trans elements (ribonucleoprotein complexes)

The splice sites and other consensus *cis* elements are recognized by the small nuclear ribonucleoprotein particles (snRNPs), serine-arginine (SR) protein complexes and heterogeneous nuclear ribonucleoprotein particles (hnRNP) of the splicing machinery. A highly organized spliceosome is formed to catalyze the excision of introns in two *trans*-esterification steps (Fig. 1B). As a first step, the 2'-hydroxyl group of the

adenosine at the branch site initiates a nucleophilic attack on the 3',5'-phosphodiester bond of the 5' splice site. Concomitant to the cleavage, the 5' terminal G of the intron is covalently attached to the branch site. In the second step, the 3' hydroxyl group of the 5' exon makes a nucleophilic attack on the phosphodiester bond at the 3' splice site. Although prevailing evidence favors that the catalytic step of pre-mRNA splicing are mediated by spliceosomal RNA, it is likely that a highly conserved protein, Prp8, which is component of the catalytic core functions as a cofactor to an RNA enzyme [6].

A

Consensus Splice Site Sequences of GT-AG and AT-AC Introns				References
Intron type	5' Splice site	Branch site	3' Splice site	
GT-AG	**GT**RAGT	CTRAY	Y**AG**	3
AT-AC	**AT**ATCCTTT	TTCCTTRACYCY	Y**AC**	20

B

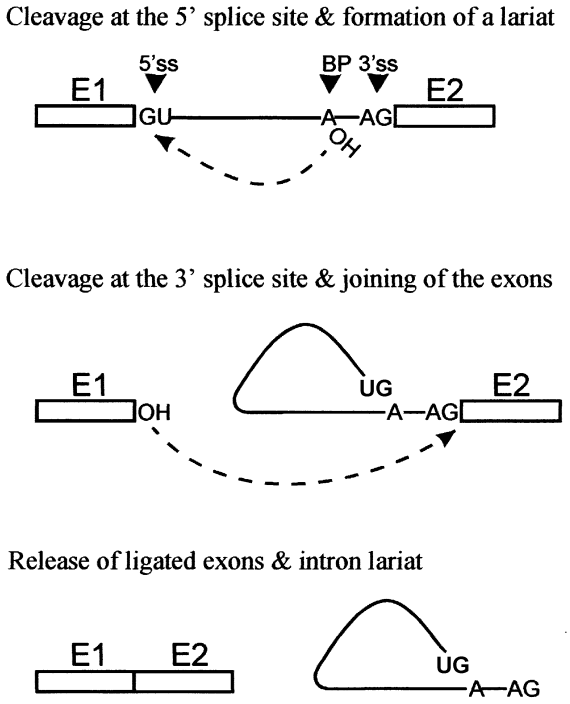

Fig. 1. (A) Comparison of the consensus characteristic *cis* elements of GT-AG and AT-AC introns. (B) Two step chemical mechanism for pre-mRNA splicing. Exons and introns are represented by rectangles and a thin line, respectively. Broken arrows indicate the nucleophilic attack. Consensus, invariant nucleotides are indicated in capital letters: R: purine (A or G), Y: pyrimidine (C or T), 5'ss: 5' splice site, 3'ss: 3' splice site, BP: branch point

1.2.1. *Small nuclear ribonucleoprotein particles (snRNPs):* The "major" spliceosome consists of five snRNPs: U1, U2, U4, U5 and U6. Each of them has a single, unique RNA component (snRNA) and several (<20) associated proteins [54]. Some of the proteins are directly involved in splicing catalysis, whereas others play a role in the formation of the accurate structure of the snRNA. The major snRNPs are present in ~10^6 copies in the nucleus [39].

1.2.2. *SR proteins:* The SR protein family is a group of essential splicing factors posessing **s**erine-a**r**ginine rich domains (SR) and one or two **R**NA-**r**ecognition **m**otifs (RRM) [18]. The SR domains are involved in protein-protein interactions, while the RRM can establish contact with *cis* elements of pre-mRNA or exposed snRNA regions.

1.2.3. *Heterogeneous nuclear ribonucleoprotein particles (hnRNPs):* Like SR proteins, also contain both RNA- and protein-protein recognition domains. More than 20 proteins constitute this family, their copy number is ~10^8 per nucleus, compared with the ~10^6 molecules of hnRNAs. Protein components of the hnRNPs bind nascent pre-mRNA immediately after transcription and can also bind to splicing activator sequences [38].

2. Discovery of the AT-AC introns and the minor spliceosome

Until 1989, all data and evidence have suggested that all pre-mRNA introns are removed by the same mechanisms, utilizing the same *cis* elements and cellular machinery. Our laboratory published the first example of introns delineated by AT and AC rather than GT and AG motifs [27]. Increasing body of evidence indicates that this kind of introns exists in a variety of organisms ranging from plants to Drosophila and vertebrates [48], albeit with a much lower frequency as compared to the major introns [32]. Usually a single minor-class introns occurs in the vicinity of major-class introns on the same pre-mRNA. The position of AT-AC intron varies within the genes and no species preference was observed for them either, but interestingly enough, members of phylogenetically related gene families harbor the AT-AC intron at conserved positions [11, 57]. This class of intron also possesses unique and highly conserved elements at the 5' and 3' splice sites and at the branch site upstream of the 3' end (Fig. 1A). An intriguing and probably significant difference between the two intron classes is that minor-class introns have no characteristic polypyrimidine tract in the neighborhood of the branch site [20]. Because the *cis* elements of the AT-AC introns differ from those of the major class, a unique set of minor snRNPs (U11, U12, U4atac, and U6atac) have evolved for the splicing of the minor-class introns [40]. The "minor group snRNPs" are found only in ~10^3 copies in the nucleus [39]. Minor snRNPs play similar roles in the splice site selection and catalytic process as their major counterparts [50, 51]. Furthermore, U5 snRNP is a common snRNP constituent in both types of spliceosome [50]. According to recent

studies, there is more overlap between the two systems at the level of the protein components. Identical splicing factors [54], splicing-associated proteins [34] and splicing-associated SR proteins [21] play crucial role in the removal of both types of intron. The AT-AC terminal dinucleotides were considered as the most characteristic feature of the minor-class introns, but recent thorough analyses showed that some introns with GT-AG terminal can also be found as minor-class introns and spliced by the U12 spliceosome. These findings revealed that for proper classification is essential to consider the complete splice site sequence and not only the terminal dinucleotides. In the light of these facts, it was reasonable to cogitate the classification and the nomenclature of these introns [48]. Therefore, the latest literature uses U2- and U12-dependent introns instead of GT-AG or AT-AC introns [32]. It was published recently, that splicing of the minor-class introns occurs more slowly, than that of the neighboring major-class introns in the same pre-mRNA. Therefore, the removal of this intron type could be the rate-limiting step in the maturation (processing) of the given pre-mRNA. Therefore, it is reasonable to assume that the U12-dependent splicing machinery can be a target in the post-transcriptional regulation of gene expression [44].

3. Spliceosome assembly

The assembly of spliceosomes on the pre-mRNA template is a well-organized process and the different snRNP components enter into the complex in a coordinated manner (Fig. 2). As a *first step*, U1 snRNP base pairs via its RNA part with the 5' splice site [35], while the U2AF protein binds to a C/U reach region (polypyrimidine tract) located between the branch site and the 3' splice site (**E complex**) [59]. U2AF is a heterodimer consisting of 65- and 35-kDa subunits. $U2AF^{65}$ interacts with the polypyrimidine tract directly, whereas $U2AF^{35}$ contacts the AG dinucleotide at the 3' splice site [19]. During the *second step,* U2AF helps U2 snRNP to associate with the branch site (**A complex**) and soon after leaves the spliceosome. In the *third step,* U4/U6 and U5 snRNPs enter the complex and form a catalytic spliceosome after conformational rearrangements of the snRNPs (**B complex**). Subsequently, cleavage takes place at the 5' splice site and a lariat is formed at the branch site. At this stage, U5 snRNP keeps the two exons in proximity in the late splicing complex and promotes cleavage at the 3' splice site and the 2 exons are joined. (**C complex**). After elimination of the intron and joining of the two neighboring exons, the spliceosome is disassembled. The former intronic region (lariat) will undergo debranching and degradation by special nucleases.

Although the characterization of the AT-AC splicing machinery has only a short history, a huge amount of information has already been accumulated. On the basis of available data, we see that despite remarkable differences in the constituents of the two heterogeneous spliceosomes, the AT-AC spliceosome assembly and catalytic steps operate according to a common principle including 2 *trans*-esterification steps [52]. While the catalytic steps do not require energy, the rearrangements within the spliceosome are ATP dependent.

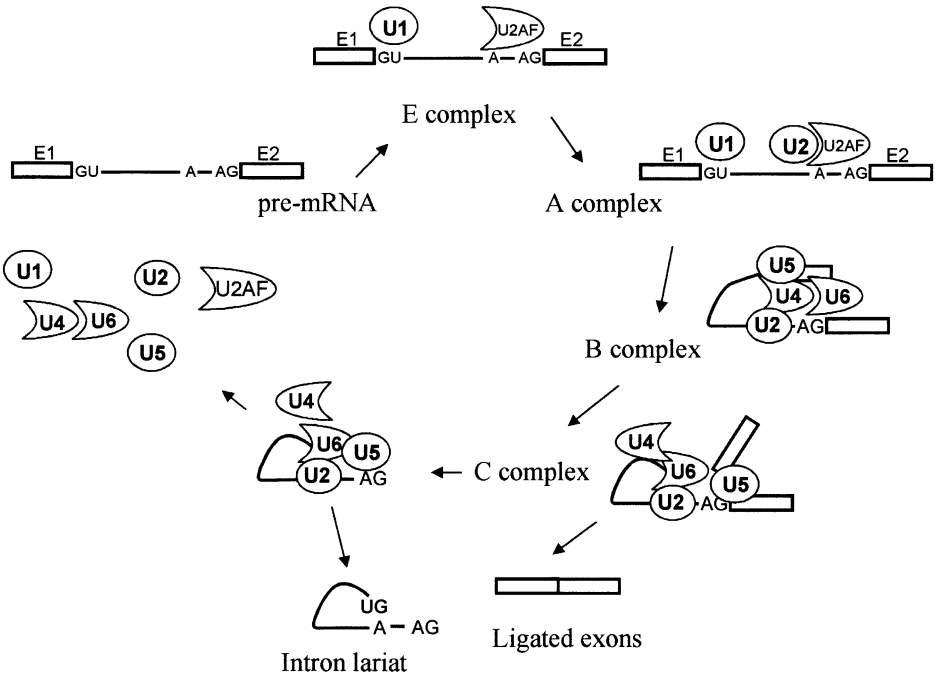

Fig. 2. Schematic representation of the spliceosome assembly and action. Pre-mRNA, containing two exons separated by an intron assembles into splicing complexes together with snRNPs. The individual snRNPs are indicated by U1, U2, U4, U5 and U6, U2AF: U2 auxiliary protein factor

4. Finding the real splice sites and pairing them

How are the true splice sites recognized, and how are the designated sites paired with their correct partners? It is obvious that short, weakly conserved *cis* sequences (splice site, branch site, polypyrimidine tract) are essential but – certainly – not sufficient for proper splice site recognition.

4.1. Identification of proper cis *elements*

One explanation for the first part of the above problem is that a given single *cis* element cooperates with another one, and in this way, they mutually increase their, otherwise weak, specificity. This kind of approach creates the basis for the so-called exon definition model [2]. Before going into details, it is advisable to dissect the problem on the basis of the actual exon positions (e.g. 5' terminal, internal or 3' terminal exon).

4.1.1. *The first exon (5' terminal exon):* It is a relatively old observation that the 5' cap structure of pre-mRNA has a positive effect on the functioning of the splicing machinery. However, the nature of the interactions involved have been enigmatic for a long time [47]. After cloning and characterization of the nuclear cap-binding complex (CBC), it has been demonstrated that CBC is required for the early steps of spliceosome assembly. CBC plays a role in the efficient association of U1 snRNP with the 5' splice site and facilitates splicing of the cap-proximal intron, while it has a very moderate effect on distal intron removal [47] (Fig. 3a).

4.1.2. *Internal exon:* According to the exon definition model, the 5' splice site can collaborate with 3' *cis* elements of the previous (upstream) intron via special SR protein factors (Fig. 3b). The theory is supported by firm experimental data. When mutations were introduced into the downstream 5' splice site, the splicing of the upstream intron was diminished, but restoration of the 5' splice site or making it more similar to the consensus resulted in more efficient intron removal under *in vitro* conditions [21, 33].

4.1.3. *The last exon (3' terminal exon):* Polyadenylation (PA) is a process which is responsible for 3' end maturation of most mRNAs. Forming poly(A) tail is significant for the stability, transport, and translation of mRNA [29]. Recent data demonstrate that polyadenylation is a highly coordinated process and it is coupled to splicing. Intact polyadenylation signal sequences seem to positively influence spliceosome formation on the last intron [41]. Mutation of the poly-A signal constituents decreased splicing efficiency and vica versa, mutation of the polypyrimidine tracts and that of the 3' splice site significantly reduced polyadenylation efficiency [9]. Protein elements of this bridging interaction are believed to be involved in the definition of the last exon [7–9] (Fig. 3c).

According to the exon definition model, exons are defined by their 5' and 3' ends. SR proteins are involved in bridging the two sides of the actual exon. More specifically, these proteins possess the ability to interact with one of the U1 snRNP-specific proteins (U1-70K) and with U2AF[35], supplying further evidence for exon defini-

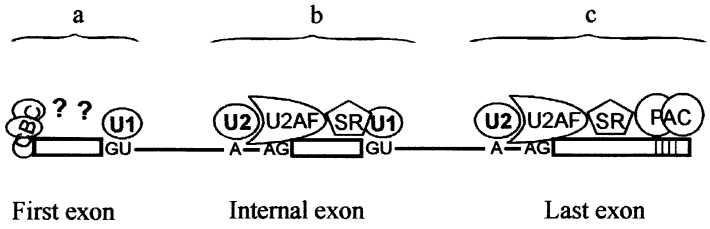

Fig. 3. Exon definition model for the spliceosome formation. Components of the macromolecular complexes formed on pre-mRNA interact across the intervening exons. CBC: cap-binding complex, U1: U1 snRNP, U2: U2 snRNP, U2AF: U2 auxiliary protein factor, SR: SR protein(s), PAC: polyadenylation complex. Detailed description is in the text

tion in the case of internal exons [15]. It is very likely that SR proteins perform a significant function in the definition of terminal exons too, but this is still yet to be established.

4.2. Splicing enhancers

Not only the *cis* elements discussed in the above paragraphs can regulate recognition of particular splice sites. Other activator sequences (enhancers), usually purine-rich regions, are located in the vicinity of the actual splice site in exonic or intronic position [17]. Splicing enhancers consist of an intricate array of *cis* elements and assemble into multiprotein complexes in adequate moments. Enhancer binding factors are SR proteins, they can promote and stabilize the binding of the U1 snRNP to a 5' splice site [5]. It has been published that ASF/SF2, PTB (polypyrimidine tract-binding protein) and different types of hnRNPs are probably constituents of a multiprotein complex assembled on splicing enhancers [4, 12, 34].

The bordering *cis* elements are slightly more conserved in U12-dependent introns than in GT-AG introns, but even those are certainly insufficient for correct definition of the splice sites (Fig. 1A). As AT-AC introns always exist in the neighborhood of GT-AG introns, it was particularly interesting to see whether they can collaborate with each other. Krainer's laboratory has shown that *in vitro* splicing of an U12-dependent intron was strongly stimulated when the 5' splice site of a downstream U2-dependent intron was situated on the pre-mRNA substrate [56]. It is still an open question, however, whether an U12-dependent intron can facilitate or not the assembly of an adjacent U2-dependent spliceosome. Using a computer prediction approach purine-rich splicing enhancers were discovered in some exons flanking U12-dependent introns, and their enhancer function was demonstrated in *in vitro* experiments [57]. SR proteins can facilitate U12-dependent splicing, but the effect is substrate (enhancer)- and SR protein-specific [22]. We identified a purine-rich splicing enhancer in close proximity to the 5' splice site in the minor-type intron of the matrilin-1 gene. After deletion of this intronic region, we found significantly lower splicing activity in *in vivo* experiments [45]. On the basis of these recent data, a picture emerges that strongly reminds us of that of the conventional U2-dependent splicing.

4.3. Correct ligation of the partner sites

Despite an appreciable amount of data presented in the previous sections, it is still not obvious, how the defined splice sites are joined. In short introns, the two splice sites can find each other directly (intron definition model). However, when the intron size is several thousand nucleotide long or even longer, it is a real challenge matching them. According to one model, transcription and splicing processes are not regarded as separate events [23]. This model presumes an intimate, direct interaction

between transcription and splicing machineries. Co-transcription/splicing practically means that the 3' part of a given gene is still involved in transcriptional events, while spliceosome assembly has already started at the 5' terminal region of the precursor mRNA [16].

Many lines of evidence suggest unexpected roles for transcription factors in linking transcription and pre-mRNA processing [37]. The largest subunit of RNA pol II was first implicated in pre-mRNA splicing and polyadenylation through its C-terminal domain (CTD) [10]. The interaction of splicing factors with phosphorylated RNA pol II CTD is probably essential for spliceosome function [43]. Moreover, polyadenylation factors and capping enzymes are also targeted onto pre-mRNA by binding to the phosphorylated CTD of RNA pol II [1]. Lai and co-workers showed that pre-mRNA splicing may be mechanistically coupled to transcription through a viral transactivator [30]. Therefore, in addition to RNA pol II and TFIID, another category of transcription factors, DNA binding transactivators, can play a role in coordinating transcription and RNA processing. Transactivators bind to a sequence-specific element nearby the promoter and primarily play a role in initiation of transcription.

Sox proteins are known as architectural transcription factors, which bind to the minor groove of the DNA via their high mobility group (HMG) domain and are implicated in tissue- and developmental-stage specific gene regulation [25]. *In vivo* transgenic mice studies proved that Sox6 together with L-Sox5 plays an essential role in chondroblast differentiation and transcriptional regulation of cartilage protein genes [49]. *In vitro* studies also confirmed that Sox6 together with L-Sox5 and Sox9 factors binds to the cartilage-specific enhancer element of the *Col2a1* gene [31]. More recently, it was published that Sox6 and Sox9 factors are involved in post-transcriptional regulation as well, namely in the formation of the spliceosome [42]. The exact function of Sox proteins is still not known in the spliceosomal complex, but it is clear that without them *in vitro* splicing reaction does not take place.

At present, a plausible mechanism is that bridging transactivators recruit essential splicing factors including non-snRNP SR proteins and spliceosomal snRNPs, thereby increasing their concentration in the vicinity of the promoter of an actively transcribed gene [30]. It remains to be examined, however, whether these kinds of transactivators can recruit SR proteins to the RNA pol II or even become incorporated into the spliceosome. More specifically, this kind of scenario (co-transcription/splicing model) would narrow the list of potential splice sites.

The so-called "packaging phenomenon" can further colour the picture. Primary RNA transcripts can be packaged into special ribonucleoprotein particles. These particles look like nucleosomes in which, instead of histones, diverse hnRNPs form the core [24]. For example, hnRNP$(C1)_3C2$ protein tetramers constitute the core unit and a ~230 nucleotides long RNA is twisted around it. Three of these tetramers can form a triangular complex containing about 700 nucleotides of RNA. After forming this kind of highly ordered structure, distal sequences (e.g. 5' and 3' ends of introns) come close and create direct liaison with each other. Biochemical data were verified by electron microscopy, visualizing higher order structures of a purified transcript packaged into hnRNP [58].

CONCLUSIONS

The excision of pre-mRNA introns is a very intricate multistep process, which involves snRNPs and auxiliary protein factors. Proper splice sites frequently differ from the consensus sequences and many cryptic sites exist in the pre-mRNA that match the consensus, but are not normally recognized by the splicing machinery. Exons are usually short stretches and their bordering elements can cooperate with each other. Conversely, the splice sites and their adjacent sequence units strongly influence their own identification. The exon definition model suggests how exons and their splice sites are initially recognized (Fig. 3). The 5' cap structure and cap binding proteins help in the identification of the 5' splice site of the first intron. In internal exons, bridging interactions define two sides of the exons (the 3' splice site of the upstream intron and the 5' splice site of the downstream intron) via bound SR proteins. In this way, otherwise weak splice sites can form a strong association to make themselves attractive for the splicing apparatus. Protein factors involved in polyadenylation process positively affect recognition of the 3' terminal exon and removal of the last intron. Splicing enhancers can further increase the probability of recognition of splice sites in tissue- and developmental stage-specific fashion. The exon definition model assumes that, first the splicing machinery recognizes a pair of splice sites around a short exon and that this process promotes spliceosome formation and intron elimination. In small introns, a pair of splice sites of a given intron can be directly recognized by the spliceosome, as interpreted by the intron definition model. Exon and intron definition models are not mutually exclusive, both of them can occur in one particular pre-mRNA depending on the exon or intron size. In addition to the identification of true splice sites, another intriguing question is how the exons are juxtaposed. The logic of the co-transcription/splicing and RNA packaging models suggests how splice sites are ligated to their correct partners. As transcription (synthesis of pre-mRNA) and splicing (spliceosome formation) proceeds in parallel, the number of feasible partners is limited. Through the formation of special "ribonucleosomes", even those splice sites can come close to each other, which are otherwise located far from each other in the precursor RNA. Although significant progress has been made in the resolution of the splice site choice problem, we are still far from the complete understanding of the whole mechanism. If we consider that the human genome has only 30,000–40,000 protein-coding genes, while the proteome (totality of proteins) has 50,000–60,000 members, it is easy to estimate the significance of the splice site choice problem, as recognition and proper cleavage steps stand behind every alternative splicing decision.

ACKNOWLEDGEMENTS

We are grateful to Professors Ferenc Solymosy and Pál Venetianer for critical reading of the manuscript. This work was supported by grants PECO CIPDCT940022, OTKA T022224, T029142 and T034399 to I.K. and OTKA F022140 to T.R.

REFERENCES

1. Barksdale, S. K., Baker, C. C. (1995) Differentiation specific alternative splicing of bovine papillomavirus late mRNAs. *J. Virol. 69*, 6553–6556.
2. Berget, S. M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem. 270,* 2411–2414.
3. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA 75,* 4853–4857.
4. Chen, C. D., Kobayashi, R., Helfman, D. M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev. 13,* 593–606.
5. Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H., Black, D. L. (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol. Cell. 5,* 949–957.
6. Collins, C. A., Guthrie, C. (2000) The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol. 7,* 850–854.
7. Cooke, C., Alwine, J. C. (1996) The cap and the 3'-splice-site similarly affect polyadenylation efficiency. *Mol. Cell. Biol. 16,* 2579–2584.
8. Cooke, C., Alwine, J. C. (2002) Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal. *Mol. Cell. Biol. 22,* 4579–4586.
9. Cooke, C., Hans, H., Alwine, J. C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol. Cell. Biol. 19,* 4971–4979.
10. Dantonel, J. C., Murthy, K. G., Manley, J. L., Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature 389,* 399–402.
11. Deak, F., Wagener, R., Kiss, I., Paulsson, M. (1999) The matrilins: a novel family of oligomeric extracellular matrix proteins. *Matrix Biol. 18,* 55–64.
12. Dirksen, W. P., Li, X., Mayeda, A., Krainer, A. R., Rottman, F. M. (2000) Mapping the SF2/ASF binding sites in the bovine growth hormone exonic splicing enhancer. *J. Biol. Chem. 275,* 29170–29177.
13. Dreyfuss, G., Kim, V. N., Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell. Biol. 3,* 195–200.
14. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., Burd, C. G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem. 62,* 289–321.
15. Fu, X. D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA 1,* 663–680.
16. Goldstrohm, A. C., Greenleaf, A. L., Garcia-Blanco, M. A. (2001) Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene 277,* 31–47.
17. Grabowski, P. J., Black, D. L. (2001) Alternative RNA splicing in the nervous system. *Prog. Neurobiol. 65,* 289–308.
18. Graveley, B. R. (2000) Sorting out the complexity of SR protein functions. *RNA 6,* 1197–1211.
19. Graveley, B. R., Hertel, K. J., Maniatis, T. (2001) The role of $U2AF^{35}$ and $U2AF^{65}$ in enhancer-dependent splicing. *RNA 7,* 806–818.
20. Hall, S. L., Padgett, R. A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol. 239,* 357–365.
21. Hastings, M. L., Krainer, A. R. (2001) Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA 7,* 471–482.
22. Hastings, M. L., Krainer, A. R. (2001) Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell. Biol. 13,* 302–309.
23. Hirose, Y., Manley, J. L. (2000) RNA polymerase II and the integration of nuclear events. *Genes. Dev. 14,* 1415–1429.
24. Huang, M., Rech, J. E., Northington, S. J., Flicker, P. F., Mayeda, A., Krainer, A. R., Le Stourgeon, W. M. (1994) The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol. Cell. Biol. 14,* 518–533.

25. Kamachi, Y., Uchikawa, M., Kondoh, H. (2000) Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet. 16,* 182–187.
26. Kessler, O., Jiang, Y., Chasin, L. A. (1993) Order of intron removal during splicing of endogenous adenine phosphoribosyl-transferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol. 13,* 6211–6222.
27. Kiss, I., Deak, F., Holloway, R. G. Jr., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., Goetinck, P. F. (1989) Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to alpha chains of beta 2 integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. *J. Biol. Chem. 264,* 8126–8134.
28. Konarska, M. M., Padgett, R. A., Sharp, P. A. (1984) Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell 38,* 731–736.
29. Kuo, H. C., Nasim, F. H., Grabowski, P. J. (1991) Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science 251,* 1045–1050.
30. Lai, M. C., Teh, B. H., Tarn, W. Y. (1999) A human papillomavirus E2 transcriptional activator. The interactions with cellular splicing factors and potential function in pre-mRNA processing. *J. Biol. Chem. 274,* 11832–11841.
31. Lefebvre, V., Li, P., de Crombrugghe, B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J. 17,* 5718–5733.
32. Levine, A., Durbin, R. (2001) A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res. 29,* 4006–4013.
33. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., Mattaj, I. W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev. 10,* 1683–1689.
34. Lou, H., Helfman, D. M., Gagel, R. F., Berget, S. M. (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol. Cell. Biol. 19,* 78–85.
35. Madhani, H. D., Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet. 28,* 1–26.
36. Maniatis, T., Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature 416,* 499–506.
37. McCracken, S,. Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., Bentley, D. L. (1997)5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev. 11,* 3306–3318.
38. Modafferi, E. F., Black, D. L. (1999) Combinatorial control of a neuron-specific exon. *RNA 5,* 687–706.
39. Montzka, K. A., Steitz, J. A. (1988) Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci. USA 85,* 8885–8889.
40. Mount, S. M. (1996) AT-AC introns: an ATtACk on dogma. *Science 271,* 1690–1692.
41. Niwa, M., Berget, S. M. (1991) Mutation of the AAUAAA polyadenylation signal depresses *in vitro* splicing of proximal but not distal introns. *Genes Dev. 5,* 2086–2095.
42. Ohe, K., Lalli, E., Sassone-Corsi, P. (2002) A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA 99,* 1146–1151.
43. Orphanides, G., Reinberg, D. (2000) RNA polymerase II elongation through chromatin. *Nature 407,* 471–475.
44. Patel, A. A., McCarthy, M., Steitz, J. A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J. 21,* 3804–3815.
45. Rauch, T., Kiss, I. (2002) Unpublished observation.
46. Ruskin, B., Krainer, A. R., Maniatis, T., Green, M. R. (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell 38,* 317–331.

47. Schell, T., Kulozik, A. E., Hentze, M. W. (2002) Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. *Genome Biol. 3,* REVIEWS 1006.
48. Sharp, P. A., Burge, C. B. (1997) Classification of introns: U2-type or U12-type. *Cell 9,* 875–879.
49. Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J. M., Behringer, R. R., de Croumbrugghe, B., Lefebvre, V. (2001) The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell 1,* 277–290.
50. Tarn, W. Y., Steitz, J. A. (1996a) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell 84,* 801–811.
51. Tarn, W. Y., Steitz, J. A. (1996b) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science 273,* 1824–1832.
52. Tarn, W. Y., Steitz, J. A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci. 22,* 132–137.
53. Venter, J. C. et al. (2001) The sequence of the human genome. *Science 291,* 1304–1351.
54. Will, C. L., Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell. Biol. 13,* 290–301.
55. Wu, Q., Krainer, A. R. (1996) U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science 274,* 1005–1008.
56. Wu, Q., Krainer, A. R. (1998) Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. *RNA 4,* 1664–1673.
57. Wu, Q., Krainer, A. R. (1999) AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol. 19,* 3225–3236.
58. Wurtz, T., Lonnroth, A., Ovchinnikov, L., Skoglund, U., Daneholt, B. (1990) Isolation and initial characterization of a specific pre-messenger-ribonucleoprotein particle. *Proc. Natl. Acad. Sci. USA 87,* 831–835.
59. Zamore, P. D., Green, M. R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J. 10,* 207–214.