Histol Histopathol (1998) 13: 585-589

Histology and Histopathology

Invited Review

Mechanisms of pre-mRNA splicing: classical versus non-classical pathways

P.G. Zaphiropoulos
Department of Bioscience, Center for Nutrition and Toxicology, Karolinska Institute, Novum, Sweden

Summary. Expression of genetic information proceeds through two major biological events, transcription and translation. However, in eukaryotic cells, the primary transcript (pre-mRNA) is not the template that the translational apparatus scans through, in order to produce the corresponding protein. Pre-mRNAs undergo several modifications (cap site addition, poly A+ tail addition) prior to becoming mature mRNAs, with the most important one being the excision (splicing) of the intronic sequences. Yet, the mechanisms that regulate the splicing process and the generation of alternatively spliced mRNA products are still poorly understood. Moreover recent findings suggest that this process also has the capability to produce an additional set of RNA products that differ from typical mRNA molecules. In these novel RNA transcripts the order of the exons has been changed relative to genomic DNA. Furthermore, the properties of these transcripts suggest that they may represent circular RNA molecules.

Key words: Exon definition, Exon juxtaposition, Trans-splicing, Circular RNAs, Scrambled exons

Introduction

The splicing process in higher eukaryotes is characterized by the precise excision of introns that can be longer than 50,000 bases and the joining of exons that are rarely over 300 bases. This joining of exons across large introns implies that the molecular mechanisms employed allow (a) the accurate recognition of the exons in an environment of non-exonic sequences, and (b) the precise joining of the exons with the concomitant occurrence of intron release.

Exon definition

Although intron/exon boundaries contain conserved sequences, with nearly all introns starting with dinucleotides GU and ending with dinucleotides AG (Jackson, 1991), apart from the exception of a minor class of AU to AC introns (Hall and Padgett, 1996; Tarn and Steitz, 1996), these sequences, independently, are not sufficient to allow precise recognition of exons. Through the work of Berget (Berget, 1995 and references therein) it has been established that splice sites are recognized as pairs and define an exon only when there is a certain proximity between these two sites. In line with this, artificial expansion of exons beyond 500 nucleotides resulted in exon skipping, implying that the expanded exons failed to be recognized by the splicing machinery (Sterner et al., 1996).

Terminal exons, that is first and last exons require special mechanisms for their recognition that appear to involve cap and poly A+ binding proteins. Moreover last exons are in general longer than internal exons with an average size of about 600 nucleotides.

In molecular terms exon recognition is accomplished by the interactions of certain components of the spliceosomal complex with the 5' and the 3' splice sites. U1 snRNA base pairs with conserved sequences of the 5' splice site and U2 snRNA with the branch site in the vicinity of the 3' splice site consensus. Additional splicing factors, including U2AF, SR proteins and hnRNPs, are also involved in the recognition process (Green, 1991; Krämer, 1996; Manley and Tacke, 1996).

Exon juxtaposition

The exon definition event is followed by the process of intron removal and exon joining. This proceeds through a two-step transesterification mechanism. First the 2' hydroxyl group of the adenosine residue of the branch site attacks the phosphate group linking the last nucleotide of the 5' exon with the guanosine residue of the GU intronic dinucleotide, thus forming the branched intermediate (lariat formation). Subsequently the 3' hydroxyl group that has been formed at the last nucleotide of the 5' exon attacks the phosphate group that links the guanosine residue of the AG intronic dinucleotide with the first nucleotide of the 3' exon,

Offprint requests to: Dr. Peter G. Zaphiropoulos, Department of Bioscience, Center for Nutrition and Toxicology, Karolinska Institute, Novum, 141 57, Huddinge, Sweden

resulting in release of the lariat and in joining of the two exons (Fig. 1). These enzymatic steps are mediated by a catalytic center that is formed by base pair interactions (helix I) between conserved residues of U2 and U6 snRNAs (Madhani and Guthrie, 1992; Nielsen, 1994; Newman, 1997). In addition, several other components of the spliceosomal complex including SR and hnRNP proteins are also involved in the mechanisms that bring together defined exons (Berget, 1995).

Alternative patterns of splicing

An intriguing property of pre-mRNA splicing is that not all exons that are of a size that allows them to be properly defined are always used in juxtaposition events. In different tissues and during different developmental stages certain exons can be skipped or defined differently using neighboring splice sites (Breitbart et al., 1987; Horowitz and Krainer, 1994). This implies that specialized splicing factors regulating these phenomena are likely to exist. Good candidates for this regulatory process are the growing number of the SR proteins as well as the hnRNPs (Cáceres et al., 1997; Chandler et al., 1997; Cogan et al., 1997; Du et al., 1997; Gallego et al., 1997). However the fact that, under certain conditions, exclusion of well definable exons can occur, suggests that the juxtaposition events do not necessarily have to occur between neighboring exons. Moreover this finding raises the question as to whether more significant perturbations from the model of exon joining that is solely based on the linear proximity of the exons can occur. For example, can the order of the exons in spliced transcripts be different to the one present in the pre-mRNA?

Scrambled transcripts

The first report of transcripts that contained exons joined in an order that was different to that of genomic DNA appeared in 1991 (Nigro et al., 1991). Certain transcripts of the tumor suppressor DCC (Deleted in Colorectal Carcinomas) gene had exon E upstream of exon B or exon D upstream of either exon C, or exon B, or exon A, even though the order of the exons in genomic DNA is A-B-C-D-E. Moreover, the ETS oncogene was found to produce transcripts that have either exon c or exon d upstream of exon a1, even though the order of the exons in genomic DNA is a1-a2-b-c-d. (Cocquerelle et al., 1992, 1993). Furthermore, the cytochrome P450 2C24 gene was found to produce a transcript that had exon 4 upstream of exon 2, even though the order of the exons in genomic DNA is 2-3-4 (Zaphiropoulos, 1993). These scrambled transcripts were found to be significantly less abundant than their corresponding canonical mRNAs, accumulated in the cytoplasm and appeared to lack a poly A+ tail.

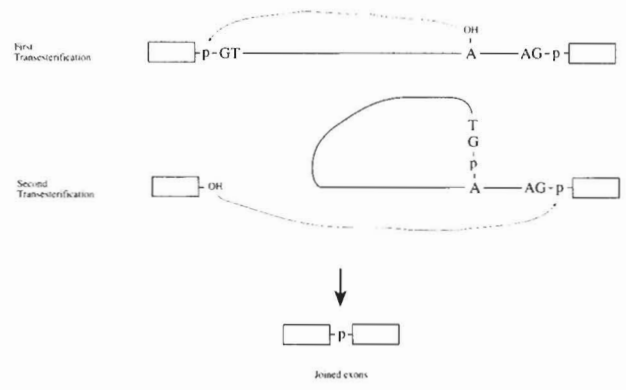

Fig. 1. Schematic diagram of the two transesterification steps involved in the splicing process. The attack of the hydroxyl groups on the phosphates is shown by arrows. Exons are indicated by boxes and introns by solid lines. The conserved GT, A and AG nucleotides are shown.

Circular RNA molecules

In 1993 the identification of an unusual transcript from the testis-determining gene Sry was reported (Capel et al., 1993). This transcript was composed of a single long exon of 1231 nucleotides that had been joined head to tail and was conclusively shown, by RNAse H analysis, to be a circular molecule. Moreover this molecule represents the most abundant Sry transcript in adult mouse testis and it is likely to be generated because of the unusual structure of the Sry genomic locus. The locus contains a small unique region of 2,7 kb (out of which the circular transcript originates) flanked by inverted repeats of over 15 kb that could form a stem-loop structure. However not only Sry but also the ETS and the cytochrome P450 2C18 genes were found to have the capability to produce RNA transcripts that are composed of single exons joined head to tail (Bailleul, 1996; Zaphiropoulos, 1997). In contrast though to Sry, these transcripts were significantly less abundant than their canonical mRNAs. Moreover in vitro, the formation of circular exons from both linear or circular pre-mRNA templates has been well documented (Schindewolf and Domdey, 1995; Braun et al., 1996; Schindewolf et al., 1996; Pasman et al., 1996).

Are the scrambled transcripts circular RNA molecules?

The low abundance of the scrambled transcripts makes technically difficult the use of direct methods to assess whether these molecules represent circular RNAs. However, the findings of a lack of both a poly A+ tail and a cap structure, and the increased resistance to exonuclease attack are consistent with these molecules being circular (Nigro et al., 1991; Cocquerelle et al., 1992, 1993; Zaphiropoulos, 1996, 1997). The alternative interpretation is that scrambled transcripts represent linear trans-spliced products that have been generated by the use of two distinct pre-mRNA templates. In vivo

there have been several reports of possible trans-splicing phenomena in higher eukaryotes (Sullivan et al., 1991; Vellard et al., 1992; Shimizu and Honjo, 1993; Romkes et al., 1993). Moreover, in vitro, splicing between distinct pre-mRNA substrates has been well documented (Bruzik and Maniatis, 1995; Chiara and Reed, 1995; Anderson and Moore, 1997). These in vitro experiments demonstrated that both the first and the second steps of the splicing process can occur between molecules that are not physically linked. Furthermore there has been direct evidence that mammalian cells can generate a trans-spliced product from the SV40 pre-mRNA (Eul et al., 1995, 1996; Graessman et al., 1996). However the properties of this molecule appear to be more similar to canonical mRNAs than to scrambled transcripts. Therefore, if pre-mRNAs have the capability to interact, under appropriate conditions, in trans, it is at least equally possible that cis interactions exploiting different combinations of exons could also occur.

Is there a correlation between circular RNA formation and exon skipping?

During the analysis of the cytochrome P450 2C24 scrambled transcript, it was found that the exons composing this molecule (exons 2, 3 and 4) had been skipped in certain mRNAs from this gene. In fact it was found that the major scrambled transcript in both liver and kidney was the exon 4-2-3-4 product while the major exon skipped mRNA produced in these tissues was the exon 1-5-6-7-8-9 product (Zaphiropoulos, 1996). The presence of the skipped exons of the mRNA into a scrambled RNA molecule prompted the formulation of the hypothesis that exon skipping and circular RNA formation might be phenomena that are mechanistically interrelated. Two interpretations rationalizing this possibility have been proposed: (a) lariats containing skipped exons undergo additional splicing events resulting in circularization of the skipped exons; (b) inverse splicing, a phenomenon known to occur in group II introns (Jarrell, 1993), provides the mechanistic rationale for the simultaneous generation of both exon skipped mRNAs and circular molecules containing the skipped exons (Fig. 2).

In order to search for correlations between exon skipping and circular RNA formation in additional genes, the expression of the cytochrome P450 2C18 was investigated in epidermis. However in contrast to 2C24, a large number of exon-skipped and circular RNAs was identified. Moreover, although the exons that had been skipped in the alternatively spliced 2C18 mRNA molecules were always found to be present in circularized RNAs, additional circular RNA species containing various combinations of exons were also identified. To rationalize this observation one can assume that, even though exon skipped and circularized RNAs might be synthesized by the same mechanism, following, at least in principle, a 1:1 ratio, differential stabilities associated with each of these molecules could result in the observed perturbation of that ratio in the

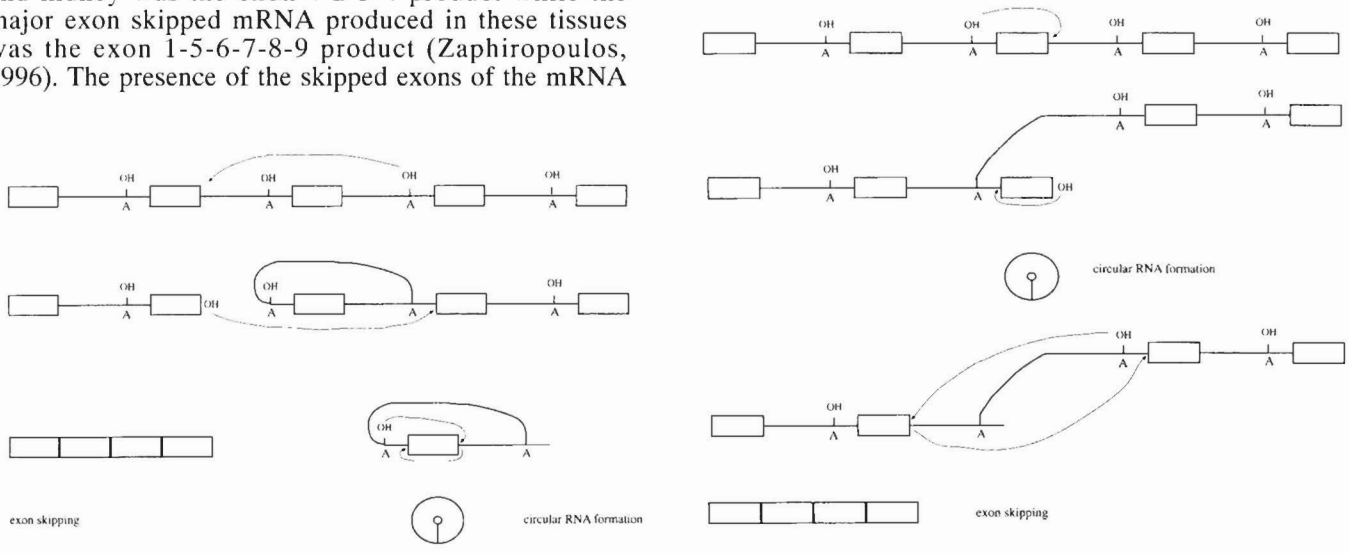

Fig. 2. Schematic diagram of additional pathways that could take place during the splicing process. **Panel A.** The 2' hydroxyl group of the branched adenosine does not attack the phosphate of the exon/intron junction immediately on its 5' side (canonical splicing) but instead the phosphates of exon/intron junctions that are located further upstream. This results in exon skipping and in generation of lariats containing exons. Additional hydroxyl attacks in these lariats result in intron release and in exon circularization. (For simplicity this panel depicts lariats containing a single exon. However, if the initial hydroxyl attacks on the pre-mRNA are on further 5' phosphates, this would result in circular RNAs containing exons that are scrambled). **Panel B.** The 2' hydroxyl group of the branched adenosine does not attack the phosphate of the exon/intron junction immediately on its 5' side (canonical splicing) but instead the phosphates of exon/intron junctions that are located downstream (inverse splicing, Jarrell, 1993). This results in the formation of a large branched intermediate containing a 3' hydroxyl group, that can attack the phosphate of its neighboring intron/exon junction resulting in exon circularization. Subsequent hydroxyl attacks of the remaining large branched molecule result in exon skipping. (For simplicity this panel depicts circularization of a single exon. However, if the initial hydroxyl attacks on the pre-mRNA are on further 3' phosphates, this would result in circular RNAs containing exons that are scrambled). Note the functional equivalence of the two processes in having the capability to generate identical spliced transcripts.

steady state. In agreement with this possibility is the absence of a 1:1 correlation between exon skipped and circularized transcripts during the expression of the ABP (Androgen Binding Protein) gene in rat testis (Zaphiropoulos, 1997).

Biological significance of circular transcripts

The finding that the splicing machinery has the capability to generate, at low frequency, a novel class of RNAs that have properties of circular molecules raises the question as to what the biological implications of this phenomenon might be. Are these molecules simply an indication of the complexities that are associated with the splicing process or do they provide clues as to as yet uncharacterized biological possibilities? It is difficult to provide clear answers to these questions. Certainly, these molecules do not necessarily keep the same open reading frame as their corresponding mRNAs, since, at the position of scrambling, joining of exons that are not in the same phase can occur. Even though circular RNA templates have the capability to promote translation, provided that ribosome entry sites are available (Chen and Sarnow, 1995) there is no direct evidence that such a potential is the major function of these circular molecules. On the other hand these species appear to accumulate in the cytoplasm suggesting that they might have a role beyond the physical constraints of the nucleus and of the splicing apparatus. Moreover, one can also suggest an evolutionary interpretation. Following the exon theory of genes proposed by Gilbert et al. (1997), it is possible that these molecules may represent remnants of a archaic capacity to join together different combinations of exonic sequences in efforts to produce functional genes. However, whether such a capacity of a primordial RNA world has any functional significance today, remains, at present, as a subject of speculation.

Acknowledgments. Funding for the experimental work that I have performed has been provided by the Karolinska Institute, the foundations Lars Hiertas Minne and Åke Wiberg, and the Swedish Natural Science Research Council.

References

Anderson K. and Moore M.J. (1997). Bimolecular exon ligation by the human spliceosome. Science 276, 1712-1716.

Bailleul B. (1996). During in vivo maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles. Nucleic Acids Res. 24, 1015-1019.

Berget S.M. (1995). Exon recognition in vertebrate splicing. J. Biol. Chem. 270, 2411-2414.

Braun S., Domdey H. and Wiebauer K. (1996). Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. Nucleic Acids Res. 24, 4152-4157.

Breitbart R.E., Andreadis A. and Nadal-Ginard B. (1987). Alternative splicing: A ubiquitous method for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. 56, 467-495.

Bruzik J.P. and Maniatis T. (1995). Enhancer-dependent interaction between 5' and 3' splice sites in trans. Proc. Natl. Acad. Sci. USA 92, 7056-7059.

Cáceres J.F., Misteli T., Screaton G.R., Spector D.L. and Krainer A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J. Cell Biol. 138, 225-238.

Capel B., Swain A., Nicolis S., Hacker A., Walter M., Koopman P., Goodfellow P. and Lovell-Badge R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73, 1019-1030.

Chandler S.H., Mayeda A., Yeakley J.M., Krainer A.R. and Fu X.-D. (1997). RNA splicing specificity determined by the coordinated action of the RNA recognition motifs in SR proteins. Proc. Natl. Acad. Sci. USA 94, 3596-3601.

Chen C.-Y. and Sarnow P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. Science 268, 415- 417.

Chiara M.D. and Reed R. (1995). A two step mechanism for 5' and 3' splice-site pairing. Nature 375, 510-513.

Cocquerelle C., Daubersies P., Majérus M.-A., Kerckaert J.-P. and Bailleul B. (1992). Splicing with inverted order of exons occurs proximal to large introns. EMBO J. 11, 1095-1098.

Cocquerelle C., Mascrez B., Hétuin D. and Bailleul B. (1993). Mis-splicing yields circular RNA molecules. FASEB J. 7, 155-160.

Cogan J.D., Prince M.A., Lekhakula S., Bundey S., Futrakul A., McCarthy E.M.S. and Phillips J.A. III. (1997). A novel mechanism of aberrant pre-mRNA splicing in humans. Hum. Mol. Genet. 6, 909-912.

Du K., Peng Y., Greenbaum L.E., Haber B.A. and Taub R. (1997). HRS/SRp40-mediated inclusion of the fibronectin EIIIB exon, a possible cause of increased EIIIB expression in proliferating liver. Mol. Cell. Biol. 17, 4096-4104.

Eul J., Graessmann M. and Graessmann A. (1995). Experimental evidence for RNA trans-splicing in mammalian cells. EMBO J. 14, 3226-3235.

Eul J., Graessmann M. and Graessmann A. (1996). Trans-splicing and alternative-tandem-cis-splicing: Two ways by which mammalian cells generate a truncated SV40 T antigen. Nucleic Acids Res. 24, 1653-1661.

Gallego M.E., Gattoni R., Stevenin J., Marie J. and Expert-Bezancon A. (1997). The SR splicing factors ASF/SF2 ad SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. EMBO J. 16, 1772-1784.

Gilbert W., De Souza S.J. and Long M. (1997). Origin of genes. Proc. Natl. Acad. Sci. USA 94, 7698-7703.

Graessmann M., Eul J., Berg B., Zimmermann C. and Graessmann A. (1996). SV40 T1 mRNA trans-splicing and translation requires that the in vitro synthesized cRNA is capped before microinjection. FEBS Lett. 394, 233-236.

Green M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell. Biol. 7, 559-599.

Hall S.L. and Padgett R.A. (1996). Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science 271, 1716-1718.

Horowitz D.S. and Krainer A.R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. Trends Genet. 10, 100-106.

Jackson I.J. (1991). A reappraisal of non-consensus mRNA splice sites.

Nucleic Acids Res. 14, 3795-3798.

Jarrell K.A. (1993). Inverse splicing of a group II intron. Proc. Natl. Acad. Sci. USA 90, 8624-8627.

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

Madhani H.D. and Guthrie C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell 71, 803-817.

Manley J.L. and Tacke R. (1996). SR proteins and splicing control. Genes Dev. 10, 1569-1579.

Newman A. (1997). RNA splicing: Out of the loop. Curr. Biol. 7, R418-R420.

Nielsen T.W. (1994). RNA-RNA interactions in the spliceosome: Unraveling the ties that bind. Cell 78, 1-4.

Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M., Oliner J.D., Kinzler K.W. and Vogelstein B. (1991). Scrambled exons. Cell 64, 607-613.

Pasman Z., Been M.D. and Garcia-Blanco M.A. (1996). Exon circularization in mammalian nuclear extracts. RNA 2, 603-610.

Romkes M., Faletto M.B., Blaisdell J.A., Raucy J.L. and Goldstein J.A. (1993). Cloning and expression of complementary DNAs for multiple members of the human cytochrome P450IIC subfamily. Biochemistry 32, 1390 (Correction).

Schindewolf C.A. and Domdey H. (1995). Splicing of a circular yeast pre-mRNA in vitro. Nucleic Acids Res. 23, 1133-1139.

Schindewolf C.A., Braun S. and Domdey H. (1996). In vitro generation of a circular exon from a linear pre-mRNA transcript. Nucleic Acids Res. 24, 1260-1266.

Shimizu A. and Honjo T. (1993). Synthesis and regulation of trans-mRNA encoding the immunoglobulin e heavy chain. FASEB J. 7, 149-154.

Sterner D.A., Carlo T. and Berget S.M. (1996). Architectural limits on split genes. Proc. Natl. Acad. Sci. USA 93, 15081-15085.

Sullivan P.M., Petrusz P., Szpirer C. and Joseph D.R. (1991). Alternative processing of androgen-binding protein RNA transcripts in fetal rat liver. Identification of a transcript formed by trans-splicing. J. Biol. Chem. 266, 143-154.

Tarn W.-Y. and Steitz J.A. (1996). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell 84, 801-811.

Vellard M., Sureau A., Soret J., Martinerie C. and Perbal B. (1992). A potential splicing factors encoded by the opposite strand of the trans-spliced c-myb exon. Proc. Natl. Acad. Sci. USA 89, 2511-2515.

Zaphiropoulos P.G. (1993). Differential expression of cytochrome P450 2C24 transcripts in rat kidney and prostate. Evidence indicative of alternative and possibly trans-splicing events. Biochem. Biophys. Res. Commun. 192, 778-786.

Zaphiropoulos P.G. (1996). Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: Correlation with exon skipping. Proc. Natl. Acad. Sci. USA 93, 6536-6541.

Zaphiropoulos P.G. (1997). Exon skipping and circular RNA formation in transcripts of the human cytochrome P450 2C18 gene in epidermis and of the rat ABP gene in testis. Mol. Cell. Biol. 17, 2985-2993.