Trans-splicing in higher eukaryotes: implications for cancer development?

Peter G. Zaphiropoulos*

Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

Trans-splicing, the possibility of exons from distinct pre-mRNAs to join together, is still a concept in gene expression that is generally regarded of limited significance. However, recent work has provided evidence that in human tumors *trans*-splicing events may precede chromosomal rearrangements. In fact, it has been suggested that the *trans*-spliced molecules could act as "guides" that facilitate the genomic translocation. This perspective highlights the development of the ideas of *trans*-splicing in higher eukaryotes during the last 25 years, from a bizarre phenomenon to a biological event that is attaining stronger recognition.

Keywords: RNA, exon, evolution, gene, recombination

DEFINITION OF TERMS

Splicing: the joining of exons from primary RNAs.

Cis-splicing: the joining of exons from a primary RNA in a 5′ to 3′ order.

Exon scrambling: the joining of exons from a primary RNA in an order where 3′ exons are positioned upstream of 5′ exons.

Trans-splicing: the joining of exons from more than one primary RNA.

Exon repetition: the presence of repeats of exon(s) in an RNA.

Exon: a sequence that is retained in a processed RNA, after removal of the intervening sequences. Exons are flanked by the major GT/AG (CG) or the minor AT/AC dinucleotides (internal exon) or a single dinucleotide and the start/end of the processed RNA (terminal exon).

Spliced leader (SL) RNA: a short RNA sequence that is *trans*-spliced to many gene transcripts in certain lower organisms, including trypanosomes and nematodes.

JOINING OF EXONS FROM DISTINCT PRE-mRNAs – EVOLUTION OF THE CONCEPT OF *TRANS*-SPLICING

The earliest reports on splicing reactions between two different RNA substrates date from 1985 (Konarska et al., 1985; Solnick, 1985). In these pioneering *in vitro* experiments the efficiency of *trans*-splicing was found to be enhanced by sequence complementarity in the intronic regions of the two mRNA precursor molecules. This was followed by evidence that in the trypanosome *Trypanosoma brucei*, and the nematode *Caenorhabditis elegans* a single RNA sequence, the SL, is *trans*-spliced to many RNAs (Murphy et al., 1986; Sutton and Boothroyd, 1986; Krause and Hirsh, 1987). A few years later, the possibility that mammalian cells may actually be involved in RNA processes that include *trans*-splicing was elegantly demonstrated by Bruzik and Maniatis (1992), when the SL RNA of *C. elegans* was shown to be capable to *trans*-splice to the adenovirus exon 2 in COS cells *in vivo*. However, this proposal

was met with a lot of skepticism. In fact, it has been suggested that even if mammalian cells have this capacity, such phenomena are not really occurring (Blumenthal, 1993). Certainly, the SL type of *trans*-splicing is not apparently taking place in higher eukaryotes. On the other hand, reports that eukaryotic exons may be joined in an order that deviates from their linear arrangement in the genome have started to accumulate since the early nineties, challenging the universality of *cis*-splicing. One early observation was that the order of exons in spliced RNAs could be reversed compared to that present in genomic DNA (Nigro et al., 1991; Cocquerelle et al., 1992). These "scrambled" RNAs were found at levels significantly lower compared to the corresponding "canonical" mRNAs, were mostly cytoplasmic and appeared to lack a polyA+ tail.

Moreover, additional reports highlighted the presence of abundant polyA+ mRNAs containing repetitions of certain exons, a phenomenon that can be rationalized by a *trans*-splicing process of independent pre-mRNA molecules (Caudevilla et al., 1998; Frantz et al., 1999). Furthermore, polyA+ mRNAs generated from gene loci present on opposite strands of a chromosome have also been reported, although in some cases, the expression level of such *trans*-spliced mRNAs was found to be quite low (Dorn et al., 2001; Labrador et al., 2001; Finta and Zaphiropoulos, 2002). Additionally, *trans*-splicing was suggested to have a role in the process of interallelic complementation in *Drosophila*, as this type of splicing was shown to also occur between different alleles (Horiuchi et al., 2003).

An elegant computational strategy was employed to detect *trans*-splicing events using non-linear exon splice junction probes on expressed sequences from the GenBank. This approach revealed 178 human genes that engage in splicing processes resulting in a change of the canonical 5′–3′ exon order (Dixon et al., 2005). Further analysis suggested that complementarity of intronic sequences has a role in promoting this non-linear splicing (Dixon et al., 2007). More recently, *trans*-splicing events that are mediated through

sequence complementarity of independent transcripts have also been observed in *C. elegans* and the unicellular eukaryote *Giardia intestinalis*, which has only few *cis*-spliced introns (Fischer et al., 2008; Kamikawa et al., 2011).
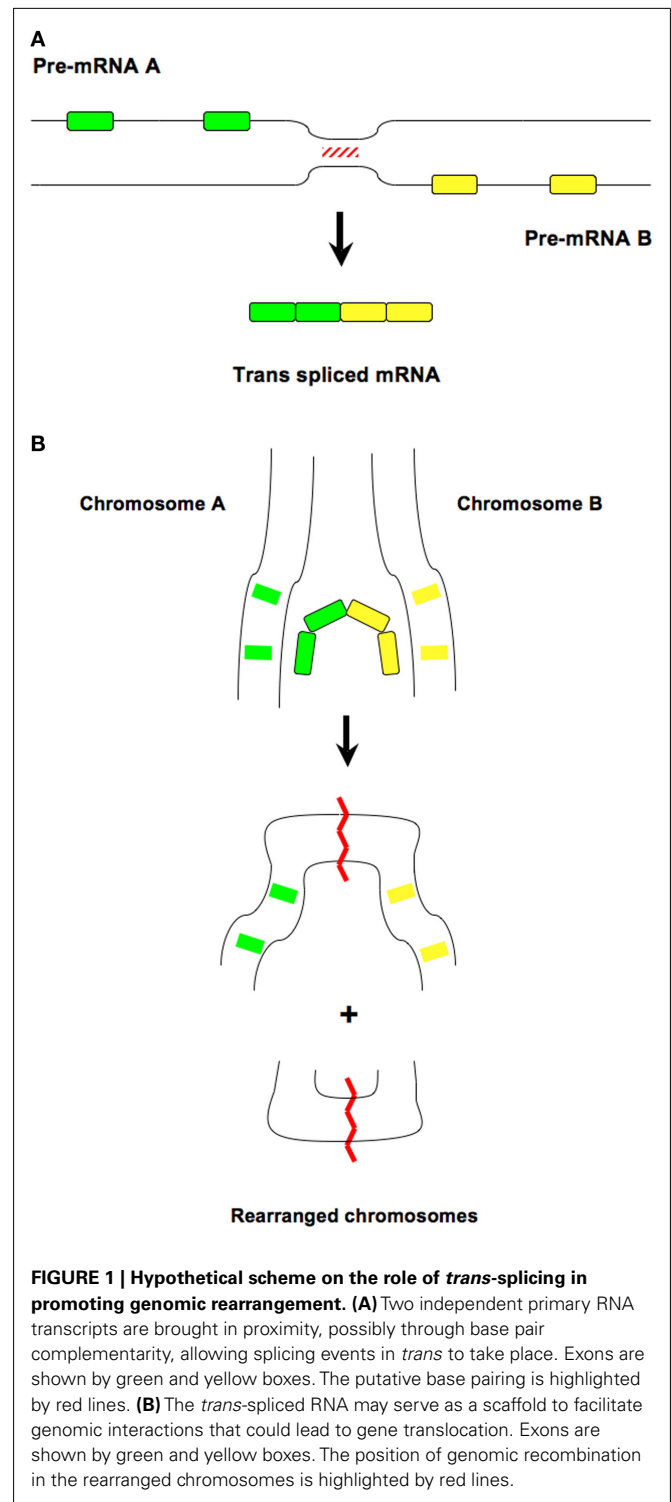
A WORD OF CAUTION

As with any novel concept the suggestion that a directional *cis*-splicing process may not account for all spliced RNAs observed in higher eukaryotic species should be subjected to rigorous quality controls. This is especially important since the methods used to detect *trans*-splicing do so in an indirect way. In a typical assay, the RNA is subjected to reverse transcription, followed by PCR amplification. However, both polymerases can engage into template switching, resulting in an artifactual generation of hybrid molecules, with this event enhanced, but not fully dependent, by repeat sequences (Zaphiropoulos, 1998; Houseley and Tollervey, 2010). This may explain a number of reports, which claimed the widespread abundance of *trans*-splicing processes, with the canonical splice signals, GT and AG not being retained (Li et al., 2009). Even RNA protection analysis, which has been used to confirm the presence of *trans*-spliced RNAs, as an independent method that avoids the involvement of reverse transcription and PCR amplification, is not full proof, unless appropriate controls are employed (Kralovicova and Vorechovsky, 2005).

One way to strengthen the argument for the finding of *bona fide trans*-spliced RNAs and not *in vitro* recombinants generated during reverse transcription/PCR, apart from the presence of canonical splice signals, is to look for premature termination codons (PTCs). Absence of PTCs in the putative *trans*-spliced molecule, and consequently maintenance of a long open reading frame, would be in line with the quality control mechanism of nonsense mediated decay (Durand and Lykke-Andersen, 2011) that *trans*-spliced mRNAs, similarly to *cis*-spliced mRNAs, are subjected to.

TRANS-SPLICING AS A TEMPLATE FOR GENE TRANSLOCATION IN ENDOMETRIAL CANCER

An unanticipated proposal on the role of *trans*-splicing in cancer biology was put forward in 2008. Namely, normal endometrial stromal cells were shown to produce a *trans*-spliced RNA, which joins the first three exons of the *JAZF1* gene on chromosome 7 to the last 15 exons of the *JJAZ1* gene on chromosome 17, that is identical to the hybrid RNA produced by the (7;17; p15;q21) chromosomal rearrangement found in endometrial stromal tumors (Li et al., 2008). This unprecedented finding raised the possibility that endometrial stromal cells capable of this *trans*-splicing may be predisposed for the genomic translocation that characterizes endometrial stromal tumors. Additionally, a possible mechanism for this translocation could be that the *trans*-spliced RNA acts as a template that facilitates the genomic fusion (**Figure 1**), a process that is in line with the RNA mediated genome rearrangement events described in ciliates (Nowacki et al., 2008). One may hypothesize that the *trans*-spliced RNA intercalates through the "breathing" DNA duplexes of the two chromosomes bringing them in proximity, and this RNA–DNA base pairing promotes strand breaks and chromosome translocation. Alternatively, the capacity of transcripts from different chromosomes to *trans*-splice may be

FIGURE 1 | Hypothetical scheme on the role of *trans*-splicing in promoting genomic rearrangement. (A) Two independent primary RNA transcripts are brought in proximity, possibly through base pair complementarity, allowing splicing events in *trans* to take place. Exons are shown by green and yellow boxes. The putative base pairing is highlighted by red lines. **(B)** The *trans*-spliced RNA may serve as a scaffold to facilitate genomic interactions that could lead to gene translocation. Exons are shown by green and yellow boxes. The position of genomic recombination in the rearranged chromosomes is highlighted by red lines.

the result of chromosomal interactions that could be mediated, at least in part, through sequence complementarity, and this proximity of the gene loci may also enhance genomic rearrangements. As a way to discriminate between these alternatives, determining the impact of exogenously added, hybrid RNA in promoting chromosomal translocations would be helpful.

In an additional reported case, a fusion transcript composed of the *SLC45A3* exon 1 joined to the *ELK4* exon 2 was found to be expressed in both benign prostate tissue and prostate cancer. Moreover, the levels of the fusion transcript did not correlate with alterations at the chromosomal level, raising the possibility that a mechanism for generation of the fusion transcript may be *trans*-splicing (Rickman et al., 2009). However, as both *SLC45A3* and the *ELK4* genes are positioned within 30 kb and in the same orientation on chromosome 1, an alternative interpretation is the presence of an extended bicistronic primary RNA out of which the "fusion" transcript is processed. Interestingly, more than a decade ago exon scrambling events on the *AML* (*MLL*) gene, which is frequently rearranged in human leukemias, that could only be partly interpreted as the result of genome duplications, were reported (Caligiuri et al., 1996; Caldas et al., 1998). Thus, the possibility of non-linear splicing processes that mimic genomic alterations had already been raised.

TRANS-SPLICING VERSUS ALTERNATIVE SPLICING – PARALLEL PATHS

Apart from certain well-documented cases of abundant *trans*-splicing events with functional implications (Gingeras, 2009), most reports on *trans*-splicing or exon scrambling are indicative of an infrequent process, with its biological significance being questioned. It is therefore possible that the majority of these non-*cis*-splicing events are products of an error prone RNA processing machinery, with limited functional consequences. However, such arguments are reminiscent of the evolution of the concept of alternative *cis*-splicing. Since its identification in the late seventies, alternative *cis*-splicing was thought for years to represent an oddity in gene expression. It is not until the last decade that it has clearly been demonstrated that this phenomenon characterizes almost all human genes (Pan et al., 2008; Wang et al., 2008). The deeper the extent of the transcriptome analysis, the higher the diversity of the identified alternative transcripts. It is therefore envisioned that with the advent of global deep sequencing technologies, which could directly sequence long RNAs at a single molecule level (Ozsolak et al., 2009), it may be possible to get convincing evidence on the pervasiveness of *trans*-splicing or exon scrambling, and their possible biological significance. Thus, the approximate 1% of human genes that engage in non-linear exon splicing, deduced from GenBank entries a few years ago (Dixon et al., 2005), is anticipated to increase. In line with this goal has been the effort to use pair end sequencing in *Drosophila* mRNAs, which identified 80 novel cases of *trans*-splicing between homologous alleles (McManus et al., 2010).

EVOLUTIONARY "TINKERING"

The concept of "bricoleur" by Jacob (2001) may be quite relevant in envisioning the biological implications of *trans*-splicing. An organism is likely to take advantage of all available "tools" in order to adapt to a constantly changing environment. As processed RNAs are composed of joined exons, more complex "tools" may be produced by a combinatorial use of exons that originate from two or more gene loci, providing a new means for expanding the diversity of the transcriptome and the proteome. Thus, similarly to alternative *cis*-splicing, which has been demonstrated to be more pronounced in higher than lower eukaryotes (Kim et al., 2007), *trans*-splicing may be a way for eukaryotic cells to take advantage of novel exon combinations that are not limited by a linear *cis* arrangement in the genome. Considering that more complex organisms do not differ so much from simpler ones, as far as the numbers of protein coding genes are concerned, it may be that significance attention should focus on the regulation of gene expression. Consequently, *trans*-splicing is to be regarded as a regulatory process that diversifies the output of exon containing genes.

FINAL NOTE

An elegant hypothesis on the evolutionary role of *trans*-splicing has recently been put forward by Blumenthal (2011). In the unicellular parasite *Giardia*, three convincing cases of *trans*-splicing mediated by base pair interactions of independent transcripts were reported (Kamikawa et al., 2011; Nageshan et al., 2011), resulting in the formation of mature mRNAs for heat shock protein 90 and dynein molecular motor protein β, which, in other organisms, are produced from single, *cis*-spliced gene loci. Thus, it is proposed that during evolution *trans*-spliced molecules, such as the ones described in *Giardia*, may have guided genomic rearrangements resulting in the formation of contiguous genes. This possibility is in line with the RNA mediated genomic rearrangement that occurs in the ciliate *Oxytricha* and the one suggested for the *JAZF1-JJAZ1 trans*-spliced RNA in endometrial cancer (Li et al., 2008; Nowacki et al., 2008). Further analysis of the genome/transcriptome of other diplomonads and related organisms is anticipated to provide additional clues in this direction.

ACKNOWLEDGMENTS

REFERENCES

Blumenthal, T. (1993). Mammalian cells can trans-splice. But do they? *Bioessays* 15, 347–348.

Blumenthal, T. (2011). Split genes: another surprise from *Giardia*. *Curr. Biol.* 21, R162–R163.

Bruzik, J. P., and Maniatis, T. (1992). Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. *Nature* 360, 692–695.

Caldas, C., So, C. W., MacGregor, A., Ford, A. M., McDonald, B., Chan, L. C., and Wiedemann, L. M. (1998). Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. *Gene* 208, 167–176.

Caligiuri, M. A., Strout, M. P., Schichman, S. A., Mrózek, K., Arthur, D. C., Herzig, G. P., Baer, M. R., Schiffer, C. A., Heinonen, K.,

Knuutila, S., Nousiainen, T., Ruutu, T., Block, A. W., Schulman, P., Pedersen-Bjergaard, J., Croce, C. M., and Bloomfield, C. D. (1996). Partial tandem duplication of ALL1 as a recurrent molecular defect in acute myeloid leukemia with trisomy 11. *Cancer Res.* 56, 1418–1425.

Caudevilla, C., Serra, D., Miliar, A., Codony, C., Asins, G., Bach, M., and Hegardt, F. G. (1998). Natural trans-splicing in carnitine

octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12185–12190.

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.

Dixon, R. J., Eperon, I. C., Hall, L., and Samani, N. J. (2005). A genome-wide survey demonstrates widespread non-linear mRNA in

expressed sequences from multiple species. *Nucleic Acids Res.* 33, 5904–5913.

Dixon, R. J., Eperon, I. C., and Samani, N. J. (2007). Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression. *Bioinformatics* 23, 150–155.

Dorn, R., Reuter, G., and Loewendorf, A. (2001). Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* 98, 9724–9729.

Durand, S., and Lykke-Andersen, J. (2011). SnapShot: nonsense-mediated mRNA decay. *Cell* 145, 324–324.e2.

Finta, C., and Zaphiropoulos, P. G. (2002). Intergenic mRNA molecules resulting from trans-splicing. *J. Biol. Chem.* 277, 5882–5890.

Fischer, S. E., Butler, M. D., Pan, Q., and Ruvkun, G. (2008). Trans-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. *Nature* 455, 491–496.

Frantz, S. A., Thiara, A. S., Lodwick, D., Ng, L. L., Eperon, I. C., and Samani, N. J. (1999). Exon repetition in mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5400–5405.

Gingeras, T. R. (2009). Implications of chimaeric non-co-linear transcripts. *Nature* 461, 206–211.

Horiuchi, T., Giniger, E., and Aigaki, T. (2003). Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, lola. *Genes Dev.* 17, 2496–2501.

Houseley, J., and Tollervey, D. (2010). Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS ONE* 5, e12271. doi:10.1371/journal.pone.0012271

Jacob, F. (2001). Complexity and tinkering. *Ann. N. Y. Acad. Sci.* 929, 71–73.

Kamikawa, R., Inagaki, Y., Tokoro, M., Roger, A. J., and Hashimoto, T. (2011). Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing. *Curr. Biol.* 21, 311–315.

Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35, 125–131.

Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1985). Trans splicing of mRNA precursors in vitro. *Cell* 42, 165–171.

Kralovicova, J., and Vorechovsky, I. (2005). Intergenic transcripts in genes with phase I introns. *Genomics* 85, 431–440.

Krause, M., and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans. Cell* 49, 753–761.

Labrador, M., Mongelard, F., Plata-Rengifo, P., Baxter, E. M., Corces, V. G., and Gerasimova, T. I. (2001). Protein encoding by both DNA strands. *Nature* 409, 1000.

Li, H., Wang, J., Mor, G., and Sklar, J. (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* 321, 1357–1361.

Li, X., Zhao, L., Jiang, H., and Wang, W. (2009). Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* 68, 56–65.

McManus, C. J., Duff, M. O., Eipper-Mains, J., and Graveley, B. R. (2010). Global analysis of trans-splicing in *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* 107, 12975–12979.

Murphy, W. J., Watkins, K. P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. *Cell* 47, 517–525.

Nageshan, R. K., Roy, N., Hehl, A. B., and Tatu, U. (2011). Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. *J. Biol. Chem.* 286, 7116–7122.

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.

Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T. G., and Landweber, L. F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* 451, 153–158.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M., and Milos, P. M. (2009). Direct RNA sequencing. *Nature* 461, 814–818.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.

Rickman, D. S., Pflueger, D., Moss, B., VanDoren, V. E., Chen, C. X., de la Taille, A., Kuefer, R., Tewari, A. K., Setlur, S. R., Demichelis, F., and Rubin, M. A. (2009). SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer. *Cancer Res.* 69, 2734–2738.

Solnick, D. (1985). Trans splicing of mRNA precursors. *Cell* 42, 157–164.

Sutton, R. E., and Boothroyd, J. C. (1986). Evidence for trans splicing in trypanosomes. *Cell* 47, 527–535.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.

Zaphiropoulos, P. G. (1998). Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I. Evidence supporting a copy choice mechanism. *Nucleic Acids Res.* 26, 2843–2848.