4152–4157 *Nucleic Acids Research, 1996, Vol. 24, No. 21*

Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract

Susanne Braun, Horst Domdey and Karin Wiebauer*

Institut für Biochemie, Genzentrum der Ludwig-Maximilians-Universität München, Feodor-Lynen-Straβe 25, D-81377 München, Germany

Received August 6, 1996; Revised and Accepted September 16, 1996

ABSTRACT

We have recently reported the first example of inverse splicing of a eukaryotic pre-mRNA intron using a whole cell extract from the yeast *Saccharomyces cerevisiae*. The concomitant circularization of the exon in the course of this splicing reaction gave rise to the hypothesis that the circular RNA species, which had been recently discovered *in vivo* in mammalian cells, were generated by inverse splicing. Here we report the formation of a circular exon in HeLa cell nuclear extracts by an inverse splicing reaction of the second intron of the human β-globin gene from a pre-mRNA transcript in which the two intron halves flanked an artificially fused, single exon. Our data demonstrate that the mammalian pre-mRNA splicing system has indeed an intrinsic capability of aligning splice sites in reverse order and that this alignment can be followed by a complete splicing reaction, whereby the discontinuous intron sequences are removed. Thus we propose that circular exons *in vivo* arise as a result of an inverse splicing reaction following the pairing of a 5′ splice site with an upstream 3′ splice site and that the frequency of this event is influenced by the presence and strength of other, competing splice sites.

INTRODUCTION

Nuclear pre-mRNA splicing can be viewed as the specific recognition and subsequent excision of intervening sequences (introns) from the authentic RNA copy of a protein-encoding gene in a process, that is mediated by a complex set of ribonucleoproteins and protein factors, called the spliceosome. In a tightly coupled reaction, the neighbouring ends of the remaining RNA stretches, the exons, are intermolecularly joined in linear order to yield the mature mRNA (for reviews see 1–3).

The concept of intron-directed mRNA splicing stems from the observation that the bases which form the borders of introns and exons are more highly conserved within the intronic part than in the exons. With the exception of the recently discovered minor class of introns (4–6), eukaryotic introns examined so far start almost exclusively with a 5′ GT and end with a 3′ terminal AG. In contrast, the exon nucleotides immediately adjacent to splice sites show much greater variation. Thus, it would seem that the nucleotide sequences which earmark the splice sites are contained within the introns. The recent discovery of exonic sequences that enhance the splicing efficiency of the adjacent upstream or downstream intron (7–10) suggests that the role of exon sequences in pre-mRNA splicing is only to support the process of intron recognition.

We previously reported that the splicing machinery of *Saccharomyces cerevisiae* is capable of recognizing an intron within a circularized pre-mRNA transcript. This finding was based on the demonstration that incubation of the circular pre-mRNA in a yeast whole cell extract under splicing conditions resulted in the production of the excised intron and the concomitant joining of the ends of the remaining exon sequences to yield a circular exon product (11).

The existence of circular RNA has also been demonstrated *in vivo* in mammalian cells (12–16). For example, the *Sry* mRNA transcript in mouse testis (12) apparently comprises a single exon. The authors of that study hypothesized that this unusual splicing event that led to exon circularization was facilitated by base pairing of the ends of a long pre-mRNA transcript with terminal inverted repeat sequences, leading to circularization of the initially linear transcript.

More recently we were able to show that in a yeast *in vitro* splicing system, circular exons could arise as a result of intron excision from a linear pre-mRNA transcript, in which the 3′ terminal region of the intron had been artificially placed upstream of its 5′ counterpart, from which it was separated by a single exon (17). Surprisingly, processing took place irrespective of whether the ends of the surrounding, discontinuous intron were allowed to form base pairs. Until then, the excision of intron sequences from a linear precursor RNA, which was composed of a single exon flanked by intron sequences, had only been demonstrated for self-splicing introns of groups I and II (18–20). In the case of group II introns this event had been termed ‘inverse splicing’ (19). The discovery of inverse splicing in a eukaryotic pre-mRNA splicing system like *S.cerevisiae* offered a possible explanation for the origin of circular exons in mammals, when base pairing of the flanking intronic sequences was not apparent (13–16).

*To whom correspondence should be addressed. Tel: +49 89 74017 414; Fax: +49 89 74017 448; Email: wiebauer@lmb.uni-muenchen.d

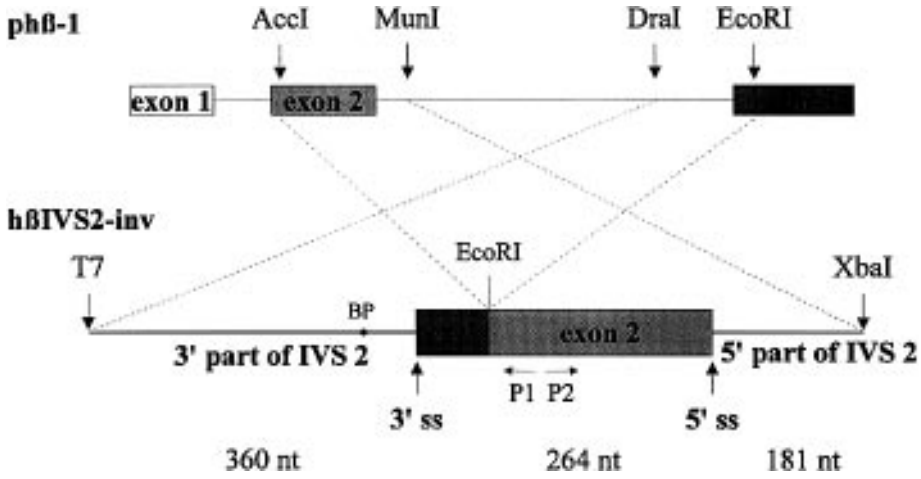

Figure 1. Construction of the DNA template for the hβIVS2-inv transcript used to study inverse splicing in a mammalian *in vitro* splicing system (for details see Materials and Methods). The intron (lines) and exon (boxes) segments were derived from the human β-globin gene (hβ). Only restriction sites relevant to the cloning procedure are shown. T7 denotes the transcription start site of the T7 RNA polymerase and *Xba*I the artificial end point. Splice sites (ss) are indicated by arrows, and the branch point (BP) by a closed circle. P1 and P2 mark the positions of the primers used for the RT–PCR reaction. The lengths of introns and exons are denoted in nucleotides (nt).

We therefore set out to investigate, whether also a mammalian pre-mRNA intron could undergo inverse splicing *in vitro* using a HeLa cell nuclear extract and a splicing substrate in which the order of splice sites of the second intron of the human β-globin pre-mRNA had been reversed. In the present study, we report the formation of a circular exon that was generated as a product of an inverse splicing reaction in nuclear splicing extracts from human cells.

MATERIALS AND METHODS

Plasmid construction

The human β-globin (hβ) gene sequences (21), used for the construction of phβIVS2-inv, were derived from the recombinant plasmid phβ-1 (22). A schematic diagram of how phβIVS2-inv was generated is shown in Figure 1. In detail, a 400 bp *Dra*I–*Eco*RI fragment (position 1101–1501 in the hβ-globin gene sequence) was ligated to a 388 bp *Acc*I–*Mun*I fragment (position 387–775) after the protruding ends of the restriction fragments had been filled in with the Klenow enzyme. In this way, the *Eco*RI site located within exon 3 was joined to the *Acc*I site located in exon 2, thereby regenerating the *Eco*RI site. Subsequently, the ligated fragments were inserted downstream of the T7 promoter between the cut back *Kpn*I- and filled in *Xba*I-restriction sites of pBluescriptIISK$^-$ (Stratagene). In the course of this procedure, the *Kpn*I site of the plasmid vector was deleted while the 3′ terminal *Xba*I site was regenerated. The resulting plasmid phβIVS2-inv, linearized with *Xba*I, served as DNA template for the *in vitro* synthesis of RNA run-off transcripts.

In vitro synthesis of RNA splicing substrates

For *in vitro* transcription, the DNA was prepared by a modification of the boiling method (23,24), omitting the RNAse step. About 50–100 ng of the linearized DNA was used without further purification. *In vitro* transcription was performed according to Melton *et al*. (25) with 5 U T7 RNA polymerase (Stratagene) in the presence of the nucleoside triphosphates ATP, CTP, UTP (500 µM each) and GTP (50 µM). Twenty µCi [α-^{32}P]GTP (New England Nuclear Corp.; 800 Ci/mmol) was added for each 10 µl reaction. m^{7}G(5′)ppp(5′)G was included in 10-fold excess over the amount of GTP to ensure that transcription started with a cap nucleotide. After 45 min at 37°C the transcripts were fractionated on 4% polyacrylamide (19:1)/8 M urea gels. The full length species was excised from the gel and the RNA was eluted, extracted with phenol/chloroform and ethanol precipitated.

Splicing reactions

Standard splicing reactions were performed in a 10 µl volume at 30°C and left overnight for convenience (10–14 h). The assays contained 1.5 mM ATP, 2 mM $MgCl_2$, 30 mM KCl, 20 mM phosphocreatine, 3% PEG 8000 and 5–10 fmol transcript (~50 000 c.p.m.). The reactions were started by the addition of 4 µl HeLa nuclear extract (7 mg/ml; Computer Cell Culture Center, Belgium). As this splicing extract had been dialysed against buffer D (20 mM HEPES–KOH, pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT) according to Dignam *et al*. (26), the final concentrations of the various components in the splicing reaction were as follows: 1.5 mM ATP, 2 mM $MgCl_2$, 70 mM KCl, 20 mM phosphocreatine, 3% PEG, 8 mM HEPES–KOH, pH 7.9, 8% glycerol, 0.08 mM EDTA, 0.2 mM DTT. Both ATP and phosphocreatine were omitted from the reaction minus ATP. For the reaction in the presence of the 2′-*O*-methyl oligoribonucleotide, $U2_{27–49}$ 5′-AUAAIAACAIAU-ACUACACUUIA-3′ (27), the oligo was added to a final concentration of 5 µM immediately before the addition of splicing extract (28). At the end of the reaction, the volume was increased to 100 µl with 50 µl 2× proteinase K buffer (200 mM Tris–HCl, pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% SDS, 50 µg/ml *E.coli* tRNA, 1 mg/ml proteinase K) and 40 µl H_2O and the incubation continued at 37°C for 30 min. Following phenol/chloroform extraction and ethanol precipitation the isolated reaction products were separated on 8% polyacrylamide (19:1)/8 M urea gels and visualized by autoradiography. Usually an overnight exposure of the dried gel to an X-ray film placed between two intensifying screens was sufficient to detect the circular splice product.

RT–PCR and DNA sequence analysis

For RT–PCR analysis, the products of three splicing reactions were separated on a denaturing 8% polyacrylamide gel, the band corresponding to the putative circular exon product was excised and the RNA eluted. One half of the isolated RNA was reverse transcribed with AMV-Reverse Transcriptase (Boehringer Mannheim) following instructions given by the supplier. One tenth of the reverse transcription reaction was used for a standard PCR. The primers used for the RT–PCR (P1: 5′-CAAAGAACCTCTG-GGTCC-3′, position 413–396; P2: 5′-CCTGATGCTGTTATGG-GC-3′, position 437–454) were located within exon 2 (Fig. 2A). The PCR products were subcloned in pBluescriptIISK$^-$ and their sequence was determined by the dideoxy method (29) using the sequencing kit from Pharmacia and primer sites within the T3 or T7 promoter region of the vector.

RESULTS

As a model for studying inverse splicing in a mammalian *in vitro* system, we used HeLa cell nuclear extracts and a splicing substrate (hβIVS2-inv) that was based on the second intron of the human β-globin gene (21). To reverse the order of splice sites, the 3′ terminal segment of the intron (plus flanking exon sequences) was placed upstream of its 5′ counterpart as shown in Figure 1.

Overnight incubation of the hβIVS2-inv transcript in a standard HeLa cell splicing extract resulted in the formation of a new RNA species which was visible as a faint band above the position of the precursor (Fig. 2B, lane 2). No such signal was observed in the absence of ATP (Fig. 2B, lane 1). This band increased in intensity over a period of 5 h (Fig. 2C, lanes 1–4), indicating that it represented a product of the *in vitro* reaction. Since this RNA species did not form in the presence of the 2′-*O*-methyl oligoribonucleotide $U2_{27–49}$, complementary to the branch site recognition region of U2 snRNA (Fig. 3), we concluded that this product resulted indeed from a pre-mRNA splicing reaction. Moreover, the fact that this product migrated abnormally in the denaturing polyacrylamide gel, marked it as a likely candidate for the expected circular exon (Fig. 2A).

To demonstrate the circular nature of the putative splice product, the corresponding band was isolated on a preparative scale from a denaturing 8% polyacrylamide gel and the eluted RNA electrophoresed once again on the same type of gel. In addition to the isolated product, a second band appeared at the position to which the linear form of the artificial exon would be expected to migrate (264 nt; Fig. 4A). This band could only be interpreted as being derived from a circular RNA species, which had been linearized by random nicking during isolation and denaturation prior to the second gel analysis. Any other molecule with abnormal migration behaviour, such as a lariat or a Y-shaped structure, would have given rise to more than one species under these conditions.

Finally, the gel-purified circular RNA was converted into double-stranded DNA by RT–PCR with primers that allowed amplification of exon sequences across the splice junctions. The resulting PCR products were cloned and sequenced. The sequence obtained with several independently derived clones provided the final evidence of the circular RNA species being a product of an inverse splicing reaction in which the ends of the single exon had been joined intramolecularly at the sites of intron excision (Fig. 4B).

DISCUSSION

The term inverse splicing refers to the excision of intron sequences from an RNA precursor in which two intron halves flank a complete, single exon (‘single exon substrate’). When inverse splicing was first discovered in transcripts containing a discontinuous self-splicing intron of groups I or II, the question arose, whether such a splicing event could also be expected to take place with a similarly composed nuclear pre-mRNA precursor, the processing of which is factor-mediated and requires the recognition of the splice sites by the components of the spliceosome (19).

This question was all the more relevant as inverse splicing yielded a circular exon, and reports of the existence of circular RNAs in mammalian cells *in vivo* had just emerged in the literature. In all reported cases, the circular RNAs seemed to have resulted from intramolecular joining of authentic (13–16) or putative (12) exon ends in an mRNA splicing reaction. In the most remarkable example of a naturally occurring circular RNA, the *Sry* transcript in the adult mouse testis, >90% of this RNA was demonstrated to be of a circular nature (12). The minor species, presumably unprocessed precursor RNA, seemed to be of the single-exon type with long intron stretches enclosing an internal exon.

Interestingly, it had been shown previously (30) that such single-exon mRNA splicing substrates were able to induce spliceosome formation, indicating that the factors involved in spliceosome assembly did also recognize splice sites on either side of an exon. This observation led to the model of ‘exon definition’ introduced by Berget and co-workers (30,31); yet, no splicing product had been reported.

Here we demonstrate that in a HeLa cell nuclear splicing extract, a circular exon can be generated from synthetic pre-mRNA, in which the 3′ and 5′ halves of the second intron of the human β-globin gene transcript flanked an artificial, composite exon. It is clear that this reaction proceeded at a very slow rate with very low efficiency which might have been one of the reasons why the circular exon products have so far escaped their detection in *in vitro* assays, when single-exon substrates were employed in mRNA splicing studies (30,32,33).

With the exception of the *Sry* transcript, which is apparently processed from a single-exon precursor RNA, all other examples of circular exons *in vivo* have only been discovered by virtue of the polymerase chain reaction. In the case of the DCC and similarly the ETS1 gene (13–15), the joining of a downstream to an upstream exon lead to the formation of a circular RNA consisting of more than one exon. For whatever reason this event took place, it would not have been expected to occur with high efficiency like in the *Sry* example, as the ‘inverse splicing’ reaction had to compete with the ‘normal’ processing pathway in which the joining of exon ends in a multi-exonic precursor mRNA displays a strict 5′–3′ polarity.

While it is still too early to conclude, it is nonetheless tempting to speculate that the low efficiency that we observed with our single-exon splicing substrate, may be a reflection of the differences in transcription of the *Sry* gene in the genital ridge during development as compared with its transcription in testis. In the genital ridge cells, in which *SRY* exerts its function, transcription starts much closer to the 3′ splice site than in testis (170 nt versus >700 nt) and the resulting transcript is not spliced inversely although it continues well beyond the following 5′ splice site (>800 nt) (34).

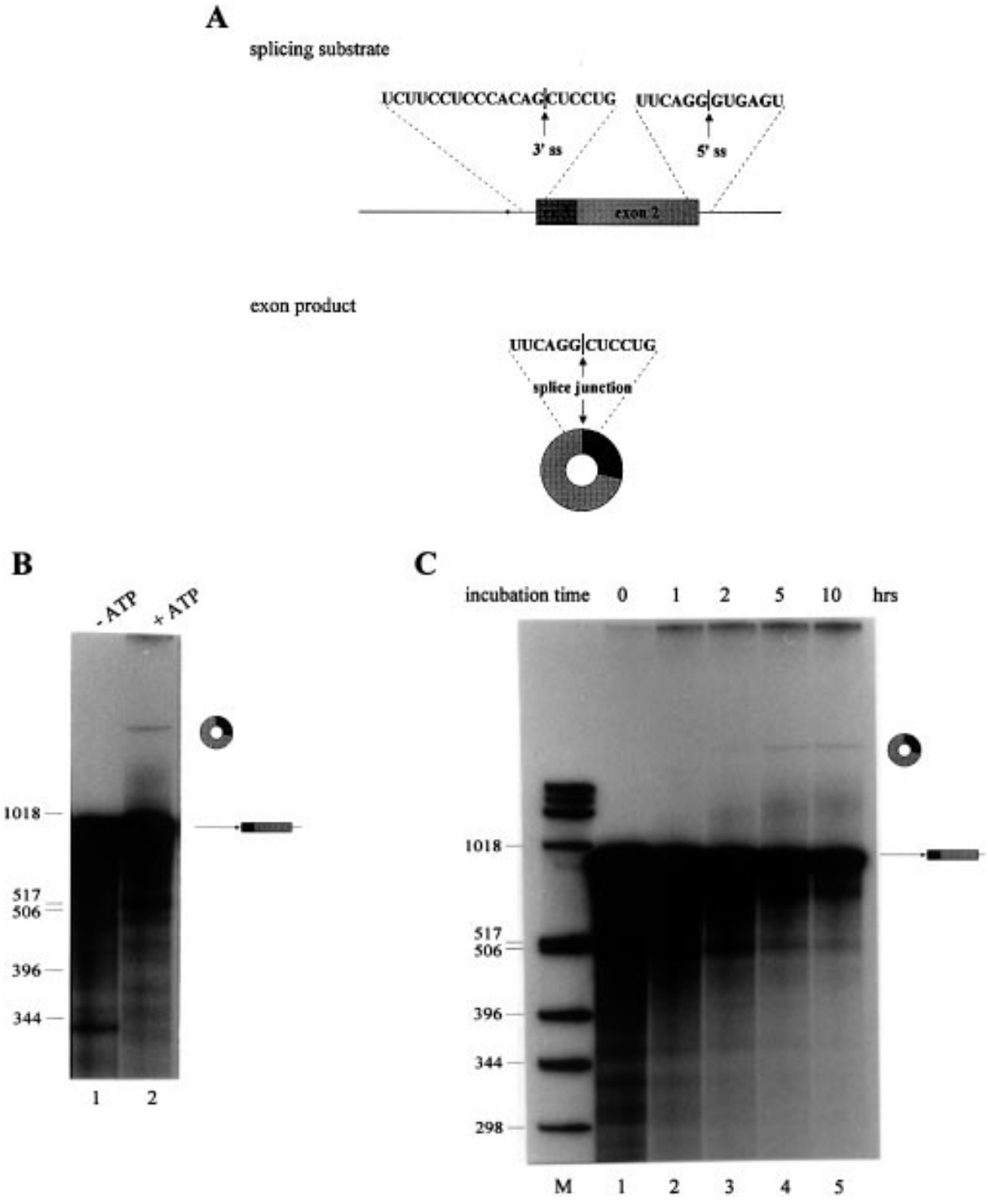

Figure 2. Fate of the hβIVS2-inv pre-mRNA upon incubation with HeLa cell splicing extract. (**A**) Structure of the RNA splicing substrate and the postulated circular exon product. Big letters indicate the base sequences surrounding the splice sites and the spliced junction of a circular exon product. (**B**) Autoradiography of the products of a standard splicing reaction with hβIVS2-inv pre-mRNA in the absence (lane 1) and presence (lane 2) of ATP. Reaction products were resolved on a denaturing 8% polyacrylamide gel. Autoradiographic exposure was overnight. (**C**) Time course of the *in vitro* reaction with the hβIVS2-inv transcript. The smear above the precursor (Fig. 2B, lane 2 and Fig. 2C, lanes 3, 4 and 5) is most probably due to non-specific polyadenylation of the pre-mRNA, as it was abolished by the addition of 1 mM cordycepin (3′-dATP; see Fig. 3 and ref. 36). Symbols on the right indicate the putative circular exon product and the precursor. M: molecular size marker, kb-ladder (Gibco), fragment lengths in nucleotides are given on the left.

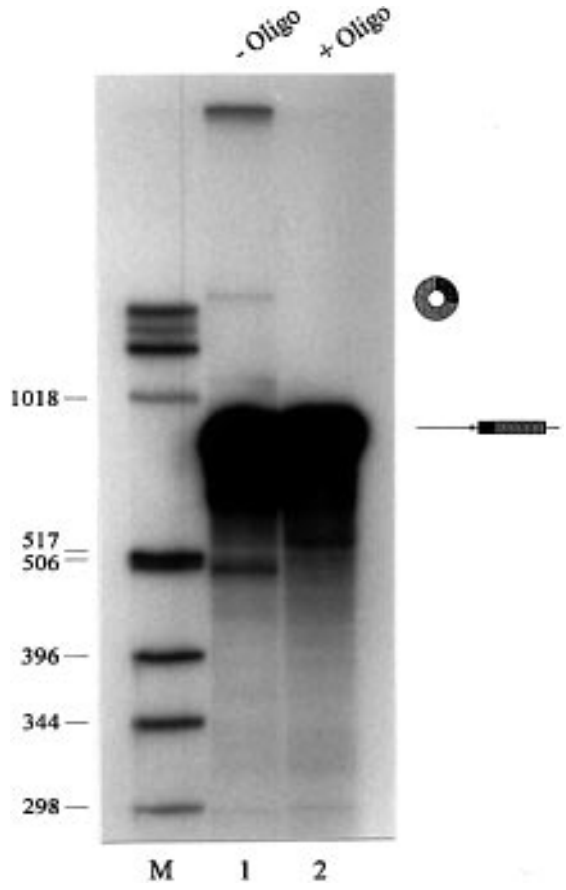

Figure 3. Inhibition of product formation in the presence of an 2′-*O*-methyl oligoribonucleotide complementary to the branch point region of U2 snRNA. hβIVS2-inv transcript was incubated overnight with HeLa splicing extract in the absence (lane 1) and presence of 5 µM (lane 2) of the antisense oligonucleotide, $U2_{27–49}$ (27). M: molecular size marker, kb-ladder (Gibco), fragment lengths in nucleotides are given on the left.

Instead, only linear RNA species could be demonstrated, possibly indicating a certain length requirement for the flanking intron sequences to be recognized as substrate for an inverse splicing reaction. However, it cannot be excluded that, as the authors suggested, circularization of the transcript would enhance the inverse splicing efficiency. We are currently pursuing this aspect of structural requirements for the efficiency of inverse splicing.

The fact that individual exons were recently also found to be circularized *in vivo*, precisely at the splice junctions of the flanking introns, albeit with very low efficiency (16), reflects an intrinsic ability of the respective splicing factors to align splice sites in either direction, in agreement with the postulated three-dimensional diffusion mechanism of splice site pairing (35). Whereas these findings apparently lend further *in vivo* support to the model of exon recognition they could also be seen as indication for the intron-directed splicing mechanism, we advocated for in our introduction. We now propose that the splicing reaction that can follow the recognition and pairing of splice sites across an exon *in vivo* and *in vitro* is a remnant of the activity of self-splicing introns and that the frequency with which it occurs in mammalian pre-mRNAs depends on the presence and strength of *cis*-competing splice sites.

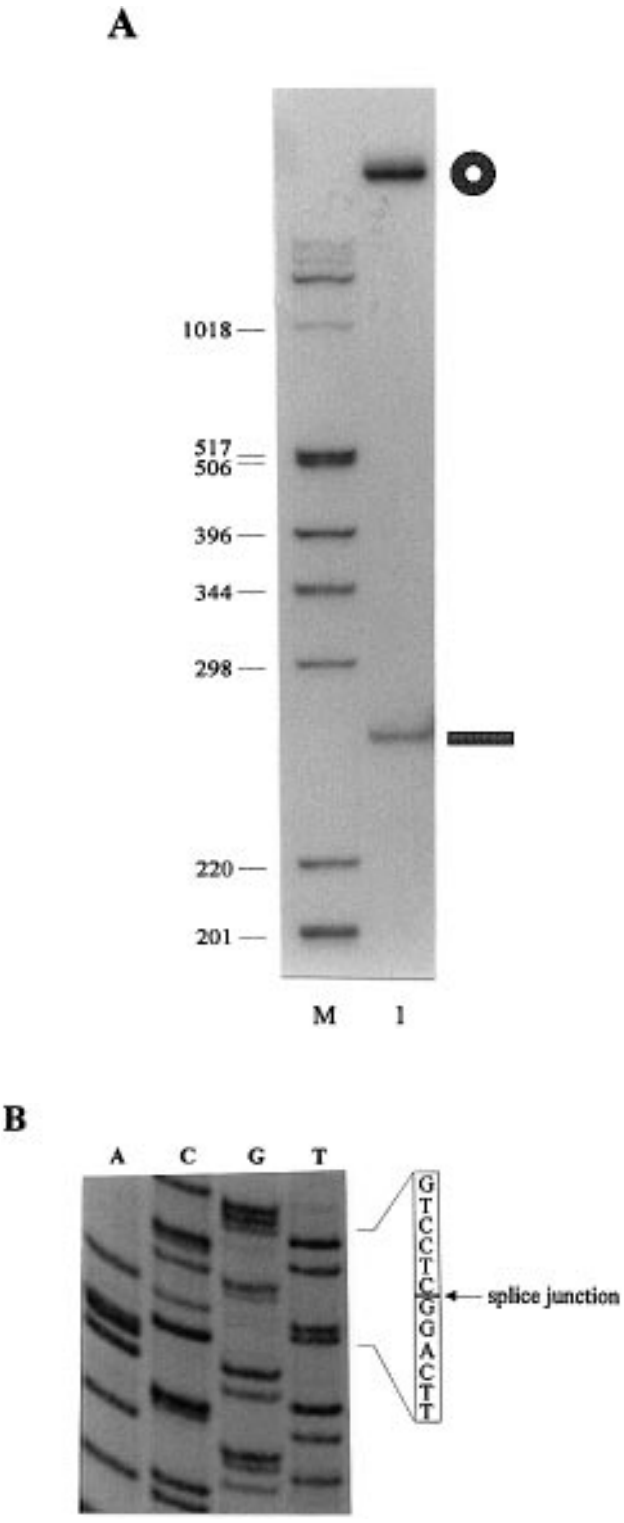

Figure 4. (**A**) Conversion of the abnormally migrating splice product into the linear form of the internal exon after isolation and re-electrophoresis. The putative circular exon was eluted from a gel, incubated at 95°C for 5 min in 80% formamide buffer and electrophoresed again on a denaturing polyacrylamide gel. Symbols on the right indicate the circular exon and its randomly linearized form. M: molecular size marker, kb-ladder (Gibco). (**B**) Sequence analysis of the cloned products of an RT–PCR reaction performed with the isolated putative circular exon. The sequence obtained at the spliced junction is indicated on the right.

ACKNOWLEDGEMENTS

We wish to thank Dr R. Chegwidden and Dr R. Tashian for valuable discussions and critical reading of the manuscript. We also would like to acknowledge H. Gassenhuber, A. Goppelt, H. Würl for their encouragement during the work and the preparation of the manuscript and especially G. Stumpf for help and critical comments on the figures. This work was supported by a grant from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie and by donations from the Fonds der Chemischen Industrie.

REFERENCES

1 Moore, M.J., Query, C.C. and Sharp, P.A. (1993) In Gestland, R.F. and Atkins, J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.
2 Green, M.R. (1986) *Annu. Rev. Genet.* **20**, 671–708.
3 Black, D.L. (1995) *RNA* **1**, 763–771.
4 Tarn, W.-Y. and Steitz, J.A. (1996) *Cell* **84**, 801–811.
5 Hall, S.L. and Padgett, R.A. (1994) *J. Mol. Biol.* **239**, 357–365.
6 Hall, S.L. and Padgett, R.A. (1996) *Science* **271**, 1716–1718.
7 Fu, X.D., Mayeda, A., Maniatis, T. and Krainer, A.R. (1992) *Proc. Natl Acad. Sci. USA* **89**, 11224–11228.
8 Tanaka, K., Watakabe, A. and Shimura, Y. (1994) *Mol. Cell Biol.* **14**, 1347–1354.
9 Tian, M. and Maniatis, T. (1994) *Genes Dev.* **8**, 1703–1712.
10 Humphrey, M.B., Bryan, J., Cooper, T.A. and Berget, S.M. (1995) *Mol. Cell Biol.* **15**, 3979–3988.
11 Schindewolf, C.A. and Domdey, H. (1995) *Nucleic Acids Res.* **23**, 1133–1139.
12 Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, P. (1993) *Cell* **73**, 1019–1030.
13 Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J., Kinzler, K. and Vogelstein, B. (1991) *Cell* **64**, 607–613.
14 Cocquerelle, C., Daubersies, P., Majerus, M.A., Kerckaert, J.P. and Bailleul, B. (1992) *EMBO J.* **11**, 1095–1098.
15 Cocquerelle, C., Mascrez, Hetuin, D. and Bailleul, B. (1993) *FASEB J.* **7**, 155–160.
16 Bailleul, B. (1996) *Nucleic Acids Res.* **24**, 1015–1019.
17 Schindewolf, C.A., Braun, S. and Domdey, H. (1996) *Nucleic Acids Res.* **24**, 1260–1266.
18 Puttaraju, M. and Been M.D. (1992) *Nucleic Acids Res.* **20**, 5357–5364.
19 Jarrell, K.A. (1993) *Proc. Natl Acad. Sci. USA* **90**, 8624–8627.
20 Ford, E. and Ares, M. (1994) *Proc. Natl Acad. Sci. USA* **91**, 3117–3121.
21 Lawn, R.M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980) *Cell* **21**, 647–651; (1993) GenBank accession number, V00499.
22 Wiebauer, K., Herrero, J.-J. and Filipowicz, W. (1988) *Mol. Cell. Biol.* **8**, 2042–2051.
23 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor University Press, Cold Spring Harbor, NY.
24 Wilmzig, M. (1985) *Trends Genet.* **1**, 158.
25 Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
26 Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
27 Lamond, A.I., Sproat, B., Ryder, U. and Hamm, J. (1989) *Cell* **58**, 383–390.
28 Dominski, Z. and Kole, R. (1994) *Mol. Cell. Biol.* **14**, 7445–7454.
29 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl Acad. Sci. USA* **74**, 5463–5467.
30 Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) *Mol. Cell. Biol.* **10**, 84–94.
31 Talerico, M. and Berget, S.M. (1994) *Mol. Cell. Biol.* **14**, 3434–3445.
32 Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) *Cell* **42**, 165–171.
33 Chiara, M.D. and Reed, R. (1995) *Nature* **375**, 510–513.
34 Jeske, Y.W.A., Bowles, J., Greenfield, A. and Koopman, P. (1995) *Nature Genet.* **10**, 480–482.
35 Pasman, Z. and Garcia-Blanco, M.A. (1996) *Nucleic Acids Res.* **24**, 1638–1645.
36 Tarn, W.-Y. and Steitz, J.A. (1995) *Genes Dev.* **8**, 2704–2717.