Cell, Vol. 64, 607–613, February 8, 1991, Copyright © 1991 by Cell Press

Scrambled Exons

Janice M. Nigro,* Kathleen R. Cho,*†
Eric R. Fearon,*‡ Scott E. Kern,*†
J. Michael Ruppert,*† Jonathan D. Oliner,*‡
Kenneth W. Kinzler,* and Bert Vogelstein*‡
* The Oncology Center
† Department of Pathology
‡ Program in Human Genetics
The Johns Hopkins University School of Medicine
Baltimore, Maryland 21231

Summary

Using a sensitive assay for RNA expression, we identified several abnormally spliced transcripts in which exons from a candidate tumor suppressor gene (*DCC*) were scrambled during the splicing process in vivo. Cloning and sequencing of PCR-amplified segments of the abnormally spliced transcripts showed that exons were joined accurately at consensus splice sites, but in an order different from that present in the primary transcript. Four scrambled transcripts were identified, each involving a different pair of exons. The scrambled transcripts were found at relatively low levels in a variety of normal and neoplastic cells of rodent and human origin, primarily in the nonpolyadenylated component of cytoplasmic RNA. These results demonstrate that the splicing process does not always pair sequential exons in the order predicted from their positions in genomic DNA, thus creating a novel type of RNA product.

Introduction

A surprising mechanism governing eukaryotic gene expression emerged with the discovery of RNA splicing in 1977 (Berget et al., 1977; Chow et al., 1977). RNA splicing was subsequently shown to occur through two consecutive transesterification steps. In the first step, the 5′ exon is cleaved from the adjacent intron with the concomitant formation of a 2′–5′ phosphodiester bond between the 5′ end of the intron and an A residue 20–50 bp upstream from the 3′ end of the intron. The second transesterification simultaneously joins the two exons and releases the intron as a lariat (Grabowski et al., 1984; Ruskin et al., 1984). For eukaryotic mRNA, the selection and ligation of appropriate pairs of exons are carried out through the recognition of specific consensus *cis* elements (Breathnach et al., 1978) by the spliceosome, a multicomponent complex consisting of protein and RNA (Brody and Abelson, 1985; Grabowski et al., 1985). The selection of the sequences to be joined is remarkably precise even in transcripts that extend for hundreds of thousands of base pairs. Alternative splicing can result in exclusion of specific exons from the final RNA product, a process that can generate several functional molecules from a single transcript in a regulated fashion (Breitbart et al., 1987).

In this report, we describe the serendipitous discovery of novel products of splicing resulting in markedly abnormal RNA molecules. Specifically, while searching for a spliced RNA transcript connecting two adjacent exons of a candidate tumor suppressor gene on chromosome 18 known as *DCC* (Fearon et al., 1990), we identified transcripts in which pairs of exons were joined accurately at consensus splice sites, but in an order different from that present in genomic DNA. Within these abnormal transcripts, the first nucleotide of an originally upstream exon was linked to the last nucleotide of an originally downstream exon. These observations have interesting implications for the fidelity of the splicing process and for the creation of unexpected RNA products.

Results

The observations described below were made during an exhaustive attempt to characterize a suppressor gene locus on chromosome 18. We assembled genomic clones encompassing 370 kb of contiguous sequences from a region of chromosome 18 suspected to contain the suppressor (Fearon et al., 1990). Each EcoRI fragment from this region was used as a hybridization probe in an attempt to identify regions of homology to other species; such homologous regions often encode expressed genes (Fearon et al., 1990). Several fragments were found that hybridized strongly to rodent DNA, but these fragments did not hybridize to cellular RNA immobilized on Northern blots.

To increase the sensitivity of the expression assay, an "exon-connection" scheme was designed, employing the polymerase chain reaction (PCR) and cDNA generated from the RNA of cell lines. The assay depended upon the identification of potential exons that were predicted from the sequences of homologous rat and human fragments. Each potential exon exhibited a conserved open reading frame (ORF) flanked by consensus splice site elements. If two such potential exons were joined by splicing into a single mRNA, we assumed it would be possible to amplify the spliced transcript from cDNA by using PCR and primers derived from the putative exon sequences. The sequences of two of the potential exons selected for this strategy are shown in Figure 1B. The expected direction of transcription *within* individual exons was apparent from the ORF and consensus splice sites. However, the orientation of the two exons with respect to one another was not known because the subcloned regions containing the putative exons and intervening intron(s) had not yet been mapped precisely; either of the two exon orientations shown in Figure 1A was possible. Two pairs of primers were therefore chosen for the PCR experiment, each pair potentially detecting a spliced transcript corresponding to one of the two orientations. Primers c and d would detect a PCR product of 233 bp ("spliced product I") if C and D represented consecutive exons in orientation I, and primers c′ and d′ would detect a PCR product ("spliced product II") of 219 bp if C and D represented consecutive

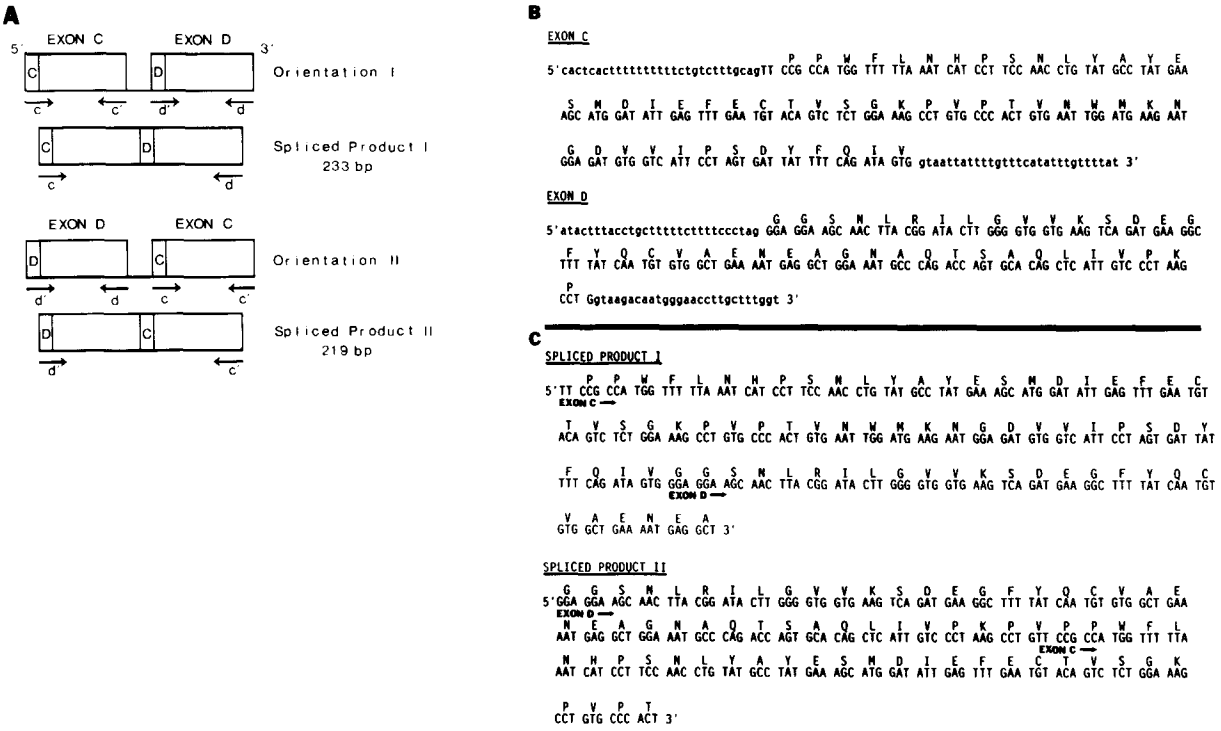

Figure 1. Normal and Scrambled Splicing of Exons C and D

(A) Orientations I and II represent the two possible arrangements of adjacent exons C and D in genomic DNA; exon C could be either upstream (orientation I) or downstream (orientation II) of exon D; an intron of undetermined length separates exon C from exon D in DNA. The 5′ end of each exon is indicated by the boxed letters, and the positions of the primers are indicated by the arrows below each exon, with the tip of the arrow pointing toward the 3′ end of the oligomer. Spliced products I and II represent PCR products generated from cDNA. Spliced product I (generated by PCR with primers c and d) would be the expected product if orientation I were present in the DNA, and spliced product II (generated by PCR with primers c′ and d′) would be the expected product if orientation II were found in DNA. Spliced products I and II are of different lengths because the primer sets used to generate them correspond to overlapping, but not identical, portions of the exons.

(B, C) Nucleotide and amino acid sequences of exons C and D, and PCR products corresponding to spliced products I and II, respectively. Intron sequences surrounding exons C and D are indicated by the lowercase letters in (B). The single letter amino acid code above the nucleotides shows the open reading frames within the exons. The first nucleotide and the direction of transcription of each exon within the spliced products are indicated by the arrows below the sequences in (C).

exons in orientation II. We expected that one of these two possible products would be found if the exons were expressed and spliced together, and that neither would be found if the exons were not expressed. To our surprise, we found both products; primers c and d produced a PCR product of the predicted size (233 bp) for spliced product I, and primers c′ and d′ produced a PCR product of the expected size (219 bp) for spliced product II. These products were initially found in PCR analyses of several human tumor cell lines (examples in Figure 2).

Cloning and sequencing of these PCR products showed that they represented precisely the spliced transcripts predicted from orientations I and II, respectively (Figure 1C). Since Southern blot hybridization experiments indicated that exons C and D were each present in single copy, there was likely to be only one orientation of the exons within genomic DNA. This suggested that one of the two products resulted from an abnormal splicing reaction. Eventual isolation of cDNA clones (Fearon et al., 1990) using exons C and D as probes suggested that orientation I was the one present in normal mRNA. Each of four cDNA clones included spliced product I and none included spliced product II. Furthermore, a long ORF was present in clones containing spliced product I, but joining of adjacent exons to spliced product II would not have preserved the open reading frame without a −1 translational frameshift between exons B and C. We concluded that spliced product I was the result of normal splicing, and spliced product II was the result of aberrant splicing in which the exons were scrambled relative to their order in genomic DNA.

We next carefully compared the level of PCR products generated from cDNA samples using primers c and d (detecting the normal spliced product I) or primers c′ and d′ (detecting the scrambled spliced product II). Control PCR assays using clones of spliced product I or II as templates produced equivalent amounts of PCR products when the same number of cycles were performed with primers c and d or with primers c′ and d′, respectively. Plasmid DNA was serially diluted, and the amount of PCR product generated from plasmid DNA was compared to that generated from cDNA of a human tumor cell line. The nor-

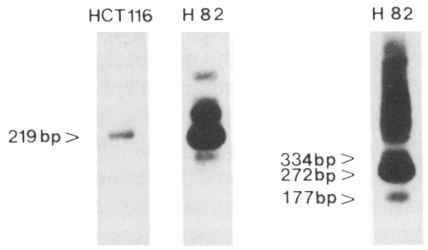

Figure 2. PCR Experiments Demonstrating Scrambled Splicing Products

RNA from the indicated cell lines was used to generate cDNA with random hexamers as primers. cDNA was used in PCR reactions using primers c′ and d′ (left two lanes) or primers b′ and d′ (right lane). Thirty-five cycles of PCR were performed (Experimental Procedures), and the PCR products were separated by electrophoresis on a 2.5% agarose gel, transferred to nylon, and hybridized with ^{32}P-labeled sequences comprising exons C and D (left two panels) or exons B and D (right panel). The 219 bp PCR product observed in the left two lanes corresponds to spliced product II in Figure 1A; in HCT116, only this product was observed, while in cell line H82, which expressed *DCC* at a higher level than HCT116, additional PCR products were observed. Numerous PCR products were also produced from the cDNA of cell line H82 using primers b′ and d′ (right lane). The four PCR products that were cloned are indicated, and their sequences are shown in Figures 1 and 4.

mal spliced product I was produced from cDNA at a level equivalent to that produced from 10^{-2} to 10^{-3} ng of plasmid DNA. Spliced product II was produced from cDNA at a level equivalent to that produced from 10^{-5} to 10^{-6} ng of plasmid DNA (Figure 3). Thus, the scrambled transcript was apparently present at approximately 1/1000th the level of the normally spliced transcript.

Using this same quantitative assay, we attempted to determine whether the scrambled transcript was confined to the nucleus. We found that the nucleus contained only

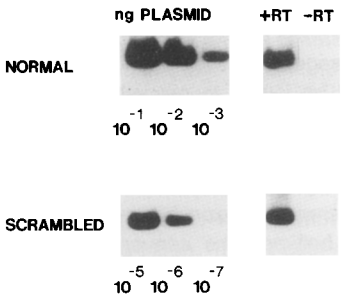

Figure 3. Relative Levels of Normal and Scrambled Transcripts Containing Exons C and D

Dilutions of plasmids representing normal (top row) or scrambled (bottom row) splicing, and cDNA generated from 3 μg of total RNA from the 577 MF cell line (+RT lane) were amplified by PCR. Enough cycles were performed so that the PCR products could be visualized on hybridization but had not reached saturation; 20 cycles were used for the top row (primers c and d; Figure 1A); and 35 cycles were used for the bottom row (primers c′ and d′, Figure 1A). The PCR products were subjected to Southern blot analysis as described in Experimental Procedures; autoradiographs of the blots are shown. The −RT lane demonstrates that no product could be amplified from a mock cDNA reaction to which all components but reverse transcriptase were added.

5% of the scrambled transcript, with the remainder present in the cytoplasm (data not shown). This localization was unlikely to be due to contamination of cytoplasmic RNA with nuclear contents, since over 98% of ribosomal transcribed spacer sequences was confined to the nucleus in the RNA preparation used. Most of the scrambled transcript was not polyadenylated, however, as 90% of it failed to bind to oligo(dT)–cellulose under conditions where actin mRNA from the same RNA preparation quantitatively bound to oligo(dT) (data not shown; see Experimental Procedures).

Additional unexpected products of splicing were observed when the PCR exon-connection scheme was used to link other potential exons to the first pair. These experiments were performed concurrently with those presented in Figure 1, before the relative orientation of exons B and D was known. From a genomic map of EcoRI fragments, we knew that putative exon C was between putative exons B and D, and we knew the sequences of the exons and the intronic regions immediately surrounding the exons. However, this information was not sufficient to predict the relative orientation of the 5′ ends of the exons with respect to one another. Two orientations were conceivable (Figure 4A). We chose primers b′ and d′ for an exon-connection experiment, guessing incorrectly that orientation II was the correct one. Using these primers in a PCR assay with cDNA as a template, a PCR product of 332 bp should have been generated if the exons were present in the DNA in orientation II; no PCR product should have been generated with primers b′ and d′ if orientation I was present in DNA and splicing had linked the exons in the conventional order. No major PCR product of the predicted size, 332 bp, was detected on a Southern blot of the PCR products using exons B, C, and D as probes. Instead, PCR products of other sizes were detected; probes corresponding to exons B and D detected several PCR products (Figure 2, right), while the probe corresponding to exon C detected no products.

The DNA fragments from some of the prominent bands were excised from the gel, subcloned, and sequenced. Three different clones were isolated: 177 bp (spliced product III), 272 bp (spliced product IV), and 334 bp (spliced product V) (Figures 4A and 4C). Eventual isolation of clones from cDNA libraries prepared by standard techniques suggested that orientation I (5′–––A–B–––C–D–E–––3′) was the one found in conventionally processed mRNA and that additional exons were found upstream of exon A, between exons B and C, and downstream of exon E. Spliced product III was apparently derived via a scrambled splicing between exons B and D, similar to that between exons C and D in Figure 1A. Spliced product IV was derived from a transcript in which exons A and B had been correctly spliced (i.e., the same orientation as in DNA), but the downstream exon D was incorrectly spliced upstream of exon A (Figure 4A). Spliced product V was derived from a transcript in which exons D and E had been correctly spliced, but the upstream exon B was incorrectly spliced downstream of exon E (Figure 4A). Thus, the order of exons A, B, D, and E had been scrambled by aberrant splicing.

The PCR assays described above were initially per-

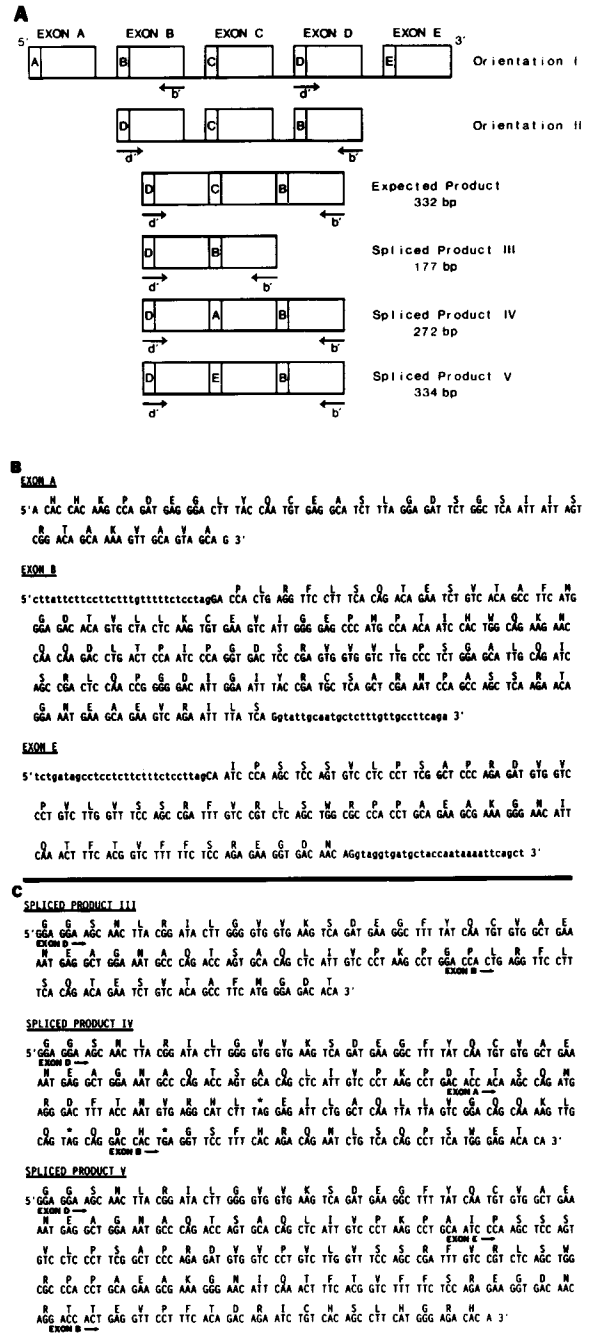

Figure 4. Additional Exons Involved in Scrambled Splicing

(A) Orientations I and II represent the possible genomic arrangements of exons B, C, and D. Orientation I depicts the order of the exons found in genomic DNA and includes two other exons (A and E), upstream and downstream, respectively, of the B–C–D group of exons. At the initiation of the experiments described, the 5′ to 3′ order of the exons was not known. Introns of undetermined length separate the exons. Primers b′ and d′ were used to generate a PCR product from cDNA. If orientation II had been the one in genomic DNA, a PCR product of 332 bp would have been obtained. No major PCR product of 332 bp was found, and instead PCR products labeled spliced products III (177 bp), IV (272 bp), and V (334 bp) were identified. The templates for these products were presumably generated through scrambled splicing, as described in the text.

(B, C) Nucleotide and amino acid sequences of exons A, B, and E and PCR products III, IV, and V, respectively. Intron sequences surrounding exons D and E are indicated by the lowercase letters in (B). The first nucleotide and the direction of transcription of each exon within the spliced products are indicated by the arrows below the sequences in (C).

formed with cDNA prepared from the RNA of tumor cell lines. It was possible that the unexpected products identified were the result of an abnormal splicing process specific to neoplastic cells. To test this hypothesis, we repeated the reactions diagrammed in Figure 1A with cDNA made from the RNA of several normal rat tissues. Rat samples were chosen because the same primers could be used (there was a high sequence conservation between the relevant sequences [Fearon et al., 1990]) and intact RNA from numerous normal tissues could be more easily obtained. The results of this experiment are shown in Figure 5A (primer pair c and d) and in Figure 5B (primer pair c′ and d′). The RNA derived from most tissues yielded the conventional spliced product I (Figure 5A). The abnormal spliced product II (Figure 5B), however, was also found in the RNA from several tissues, with brain displaying the highest levels of both normal and scrambled transcripts. Thus, scrambled splicing was not confined to tumor cells.

Discussion

We have identified abnormally spliced transcripts in which exons from a single gene have been scrambled relative to their order in genomic DNA. The abnormally spliced transcripts involved several exons and were identified in rodent and human cells and in both normal and neoplastic cell types. Splice site recognition, cleavage, and ligation occurred in all of the abnormally spliced transcripts with the usual precision of the splicing mechanism, but the process somehow failed to pair sequential exons appropriately.

The quantitation shown in Figure 3 suggested that the scrambled transcripts were present at only 1/1000th the concentration of normally spliced transcripts. However, the actual proportion of nascent RNA that undergoes scrambled splicing may be significantly higher than 0.1%. There is only one conventionally spliced transcript of the *DCC* gene, which presumably includes every exon joined in the same order as that found in DNA (ignoring alternative exon usage). However, each of the exons within the gene could be joined to any other exon by scrambled splicing, yielding a potential of over 700 different products with the 28 known exons of the *DCC* gene. We do not know whether a single processed transcript could contain more than one scrambled exon fusion. Moreover, differences in stability between the conventionally and aberrantly spliced transcripts could complicate the derived estimate of the actual proportion of primary RNA product that is abnormally spliced. We note that no scrambled exons have been observed in cDNA libraries, but the fact that most of these libraries were generated from polyadenylated RNA could create a bias against the largely nonpolyadenylated scrambled transcripts.

The factors responsible for generating the scrambled

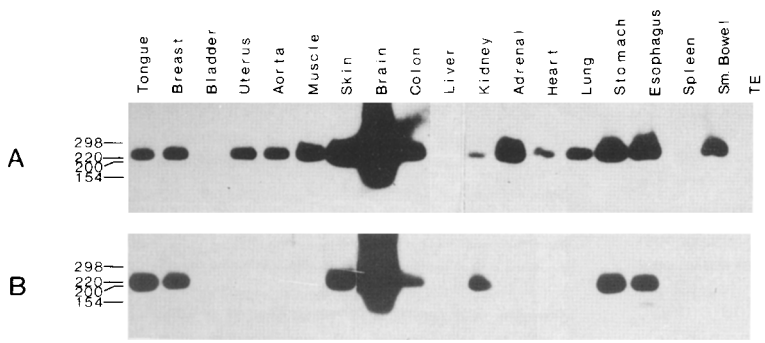

Figure 5. Normal and Scrambled Splicing in Rat Tissues

RNA from rat tissues was used to generate cDNA with random hexamers. The cDNA was used in PCR using primers c and d (A) or c′ and d′ (B). Twenty-five cycles of PCR were performed to generate the PCR products seen in (A), while the PCR products seen in (B) were the result of 35 cycles of amplification. The PCR products were separated by electrophoresis on 2.5% agarose gels, transferred to nylon, and hybridized with a ^{32}P-labeled probe comprising exons C and D. The PCR product observed in (A) corresponds to spliced product I of Figure 1A, and the PCR product observed in (B) corresponds to spliced product II of Figure 1A. The size of the coelectrophoresed markers is indicated in base pairs.

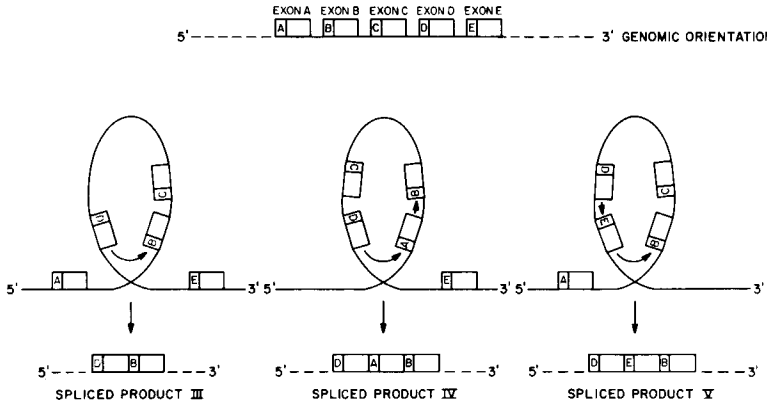

Figure 6. Possible Mechanism for Intramolecular Splicing Generating Spliced Products III, IV, and V

The genomic orientation and direction of transcription of the exons are shown at the top. The exons are separated by introns of undetermined length. RNA transcripts are represented as loops (middle). Depending on the crossover point of the loop, alternate 5′ donor splice sites are brought into close proximity to 3′ splice acceptor sites of nonsequential exons. Splicing at these sites would generate circular RNAs. Following cDNA synthesis and PCR using primers b′ and d′ (Figure 4A), the circular RNAs would give rise to spliced products III, IV, and V, shown at the bottom.

transcripts are not known, but it is reasonable to suppose that they simply represent errors in the normal splicing process. Errors by multienzyme complexes are well documented in the case of DNA replication (generating point mutations, for example; Kunkel and Alexander, 1986; Kunkel and Loeb, 1981) and in protein synthesis (miscoding amino acids; Edelmann and Gallant, 1977).

Exon scrambling could be the result of intermolecular splicing between two different *DCC* transcripts. Splicing between two transcripts (*trans*-splicing) has been observed in several systems, most notably in Trypanosomes and nematodes (Murphy et al., 1986; Sutton and Boothroyd, 1986; Krause and Hirsh, 1987; Koller et al., 1987). However, the *trans*-splicing in lower eukaryotes involves a 5′ leader with intrinsic U1-snRNP-like features (Bruzik et al., 1988; Van Doren and Hirsh 1988) and is unlikely to play a role in generating the transcripts described here. *Trans*-splicing in mammalian cells has not been observed previously in vivo, although extracts from human cells can catalyze such reactions in vitro (Konarska et al., 1985; Solnick, 1985). In vitro *trans*-splicing was greatly facilitated when the two exons contained complementary sequences within their introns. Although there was no complementarity between the intron sequences adjacent to the exons examined in our study, it is possible that such complementarity existed in more distant regions of the introns (which were not sequenced) and that these complementarities were responsible for the observed abnormal splicing patterns. Alternatively, scrambled transcripts could be the result of an intramolecular process such as that shown in Figure 6. Perhaps the physical proximity of splice sites within densely packed hnRNP particles (Choi et al., 1986) facilitates such abnormal splicing reactions.

We do not know whether scrambled splicing affects the transcripts of other genes. The only other gene we have tested in experiments analogous to those shown in Figures 1 and 4 is the *p53* gene; with *p53*, only products of the conventional orientation were identified (data not shown). We doubt, however, that exon scrambling will prove limited to the *DCC* gene. Although no previous reports have identified such products, heretofore there has been no reason to design primers that would detect their presence at low concentration.

Finally, one must ask whether there are any potential benefits associated with scrambled splicing. Because RNA can function enzymatically, it is now widely believed that RNA functioned as self-replicating molecules in the primordial environment (Reanney, 1979; Cech, 1985; Joyce, 1989). In such an RNA world, alterations of RNA must have played a central role in the exploration of new genetic functions. In modern organisms, exon recombination at the DNA level has been invoked as an efficient

means of creating new genes, and the very existence (or persistence) of exons has been hypothesized to be due to this potential (Gilbert, 1985; Darnell and Doolittle, 1986). In an RNA world, scrambled splicing could have mimicked this process, producing new variations of genes more efficiently than simple base misincorporation, as could *trans*-splicing (Sharp, 1985; Laird, 1989). One wonders whether this postulated process has continued to generate genetic diversity in more recent times. For example, it has been noted that the *src* and phospholipase C genes both contain two separate, functional domains, but the order of these functional domains is reversed within the two polypeptide chains (Stahl et al., 1988). Although this reversal could have arisen as a result of exon shuffling at the DNA level, it is conceivable that one of the genes arose through scrambled splicing of a precursor of the other, with subsequent reverse transcription and integration into the germ line. Several instances of processed and presumably reintegrated genes have been documented (Weiner et al., 1986). The scrambled splicing of RNA transcripts might thus function in evolution, even in higher eukaryotes.

Experimental Procedures

cDNA

Total RNA was prepared from frozen tissues and cell lines using guanidinium isothiocyanate (Chomczynski and Sacchi, 1987). First-strand cDNA was synthesized from 3 μg of total RNA using random hexamers as primers (Noonan and Roninson, 1988). Hexamers were removed by sodium perchlorate–isopropanol precipitation (Haymerle et al., 1986; Nigro et al., 1989). Single-stranded cDNA was used to generate cDNA libraries in λgt10 following the RNAase H protocol (Gubler and Hoffman, 1983).

PCR

For all experiments except that depicted in Figure 3, single-stranded cDNA generated from 3 μg of total RNA was amplified in a standard 50 μl PCR reaction (Saiki et al., 1988) consisting of either 25 or 35 cycles of 93°C (1 min), 55°C (1 min), and 70°C (2 min). The final concentration of $MgCl_2$ was 2 mM. The primers used were c: 5′-TTCCGCCATGGTTTTTAAATCA-3′; d: 5′-AGCCTCATTTTCAGCCACACA-3′; c′: 5′-AGTGGGCACAGGCTTTCC-3′; d′: 5′-GGAGGAAGCAACTTACGG-3′; b′: 5′-TGTGTCTCCCATGAAGGCTG-3′. The primers are derived from sequences within exons B, C, and D, respectively (Figures 1B and 4B).

Cloning and Sequencing

PCR products were separated by electrophoresis on 2.5% agarose gels. DNA bands were excised from the gel and purified by binding to glass beads (Vogelstein and Gillespie, 1979). The DNA was filled in with the Klenow fragment of DNA polymerase I, phosphorylated with T4 kinase, and ligated to EcoRI linkers. Purified, linked DNA was ligated to EcoRI-digested λgt10 arms and packaged using commercial extracts. Individual phage clones were identified by hybridization, and the EcoRI inserts were subsequently subcloned using a Bluescript vector (Stratagene Cloning Systems). T3 and T7 primers were used to sequence double-stranded DNA templates with Sequenase (Del Sal et al., 1989).

Quantitation of PCR Products

For the quantitation experiments of Figure 3, cDNA from 3 μg of RNA was amplified in a 50 μl reaction (16 mM NH_4SO_4, 67 mM Tris [pH 8.8], 6.7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 6.7 mM EDTA, 10% DMSO, 1.5 mM dNTPs, 350 ng of primers, 1 U of Taq polymerase) using cycles of 95°C (30 sec), 55°C (2 min), and 70°C (2 min). For detection of the normally spliced product, primers c and d (Figure 1) were used in a 20 cycle PCR; for the scrambled product, primers c′ and d′ were used in a 35 cycle PCR. Plasmids contained either normal *DCC* cDNA (Fearon et al., 1990) or the sequences constituting spliced product II (Figure 1). PCR products were separated on a 2% agarose gel and transferred to nylon. A ^{32}P-labeled *DCC* cDNA probe was used for detection of the normal product, and a ^{32}P-labeled oligomer spanning the junction of the scrambled product (5′-CTAAGCCTGTTCCGCCAT-3′) was used for detection of the scrambled product.

RNA Fractionation

Poly(A)$^+$ and poly(A)$^-$ RNA was isolated from total rat brain RNA using double passage over oligo(dT)–cellulose chromatography; poly(A)$^+$ RNA constituted 9% of the total RNA. To separate nuclear and cytoplasmic fractions of the human cell line 577 MF (provided by P. Andrews, Wistar Institute), cells were gently lysed in 10 mM Tris (pH 7.8), 5 mM $MgCl_2$, 5 mM DTT, 150 mM KCl, and 0.1% NP40, and nuclei were pelleted at 1000× g (Vaessen et al., 1987). RNA prepared from these nuclei constituted 3% of the total cellular RNA. Efficiency of fractionation was determined through Northern blot hybridization, using probes for actin (Hanauer et al., 1983) and 18S ribosomal transcribed spacer sequences (Gonzalez et al., 1990). Over 95% of the cellular actin mRNA was found in the poly(A)$^+$ fraction, and greater than 98% of the cellular 18S 5′ external transcribed spacer sequences was found in the nuclear RNA fraction.

Acknowledgments

This work was supported by the Clayton Fund, the McAshan Fund, the Stetler Fund, and grants GM07184, GM07309, CA43460, and CA35494 from the National Institutes of Health.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received February 8, 1990; revised November 8, 1990.

References

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA *74*, 3171–3175.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). The 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.

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

Brody, E., and Abelson, J. (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science *228*, 963–967.

Bruzik, J. P., Van Doren, K., Hirsh, D., and Steitz, J. A. (1988). *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature *335*, 559–562.

Cech, T. R. (1985). Self-splicing RNA: implications for evolution. Int. Rev. Cytol. *93*, 3–22.

Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. (1986). Heterogenous nuclear ribonucleoproteins: role in RNA splicing. *Science 231*, 1534–1539.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal. Biochem. *162*, 156–159.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell *12*, 1–8.

Darnell, J. E., and Doolittle, W. F. (1986). Speculations on the early course of evolution. Proc. Natl. Acad. Sci. USA *83*, 1271–1275.

Del Sal, G., Manfioletti, G., and Schneider, C. (1989). The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. Biotechniques *7*, 514–520.

Edelmann, P., and Gallant, J. (1977). Mistranslation in E. coli. Cell *10*, 131–137.

Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W., and Vogelstein, B. (1990). Identification of a chromosome 18q gene which is altered in colorectal cancers. Science *247*, 49–56.

Gilbert, W. (1985). Genes-in-pieces revisited. Science *228*, 823–824.

Gonzalez, I. L., Chambers, C., Gorski, J. L., Stambolian, D., Schmickel, R. D., and Sylvester, J. E. (1990). Sequence and structure correlation of human ribosomal transcribed spacers. J. Mol. Biol. *212*, 27–35.

Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell *37*, 415–427.

Grabowski, P. J., Seiler, S. R., and Sharp, P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell *42*, 345–353.

Gubler, U., and Hoffman, B. (1983). A simple and very efficient method for generating cDNA libraries. Gene *25*, 263–269.

Hanauer, A., Levin, M., Heilig, R., Daegelen, D., Kahn, A., and Mandel, J. L. (1983). Isolation and characterization of cDNA clones for human skeletal muscle α-actin. Nucl. Acids Res. *11*, 3503–3516.

Haymerle, H., Herz, J., Bressan, G. M., Frank, R., and Stanley, K. K. (1986). Efficient construction of cDNA libraries in plasmid expression vectors using an adaptor strategy. Nucl. Acids. Res. *14*, 8615–8624.

Joyce, G. F. (1989). RNA evolution and the origins of life. Nature *338*, 217–224.

Koller, B., Fromm, H., Galun, E., and Edelman, M. (1987). Evidence for in vivo *trans* splicing of pre-mRNAs in tobacco chloroplasts. Cell *48*, 111–119.

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

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

Kunkel, T. A., and Alexander, P. S. (1986). The base substitution fidelity of eucaryotic DNA polymerases. J. Biol. Chem. *261*, 160–166.

Kunkel, T. A., and Loeb, L. A. (1981). Fidelity of mammalian DNA polymerases. Science *213*, 765–767.

Laird, P. W. (1989). *Trans* splicing in trypanosomes—archaism or adaptation? Trends Genet. *5*, 204–208.

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.

Nigro, J. M., Baker, S. J., Preisinger, A. C., Jessup, J. M., Hostetter, R., Cleary, K., Bigner, S., Davidson, N., Baylin, S., Devilee, P., Glover, T., Collins, F. S., Weston, A., Modali, R., Harris, C. C., and Vogelstein, B. (1989). p53 gene mutations occur in diverse human tumour types. Nature *342*, 705–708.

Noonan, K. E., and Roninson, I. B. (1988). mRNA phenotyping by enzymatic amplification of randomly primed cDNA. Nucl. Acids Res. *16*, 10366.

Reanney, D. (1979). RNA splicing and polynucleotide evolution. Nature *277*, 598–600.

Ruskin, B., Krainer, A. R., Maniatis, T., and Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Saiki, R. K., Gelfano, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487–491.

Sharp, P. A. (1985). On the origin of RNA splicing and introns. Cell *42*, 397–400.

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

Stahl, M. L., Ferenz, C. R., Kelleher, R. L., and Knopf, J. L. (1988). Sequence similarity of phospholipase C with the non-catalytic region of src. Nature *332*, 269–271.

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

Van Doren, K., and Hirsh, D. (1988). Trans-spliced leader RNA exists as small ribonucleoprotein particles in Caenorhabditis elegans. Nature *335*, 556–559.

Vaessen, R. T., Houweling, A., and van der Eb, A. J. (1987). Post-transcriptional control of class I MHC mRNA expression in adenovirus 12-transformed cells. Science *235*, 1486–1488.

Vogelstein, B., and Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA *76*, 615–619.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. *55*, 631–661.