Activation of a cryptic 5' splice site in the upstream exon of the phage T4 *td* transcript: exon context, missplicing, and mRNA deletion in a fidelity mutant

P. Scott Chandry^{1,2} and Marlene Belfort^{1,3}

¹Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, New York 12201 USA;

²Department of Microbiology and Immunology, Albany Medical College, Albany, New York 12208 USA; ³School of Public Health Sciences of the State University of New York, New York State Department of Health, Albany, New York 12201 USA

A collection of 100 *td* mutants defective in phage T4 thymidylate synthase (TS) production was screened for splicing impairments. Splicing-defective mutants were identified by a rapid assay developed to detect imbalances in the *td* protein products (TS, the exon ligation product, and NH₂TS, encoded by the pre-mRNA). Thirteen selected mutants, confirmed to be splicing defective by an RNA-oligodeoxynucleotide hybridization assay, were all shown to be inhibited in the first step of the group I splicing pathway, cleavage at the 5' splice site. Of these, only one, SC99, appeared to be a specificity mutant. Whereas the 12 other mutants had sequence changes within the functionally important 5' and 3' domains of the intron, SC99 was shown to be an exon mutant. The G → A change at residue -3 of the upstream exon of SC99 resulted in loss of normal 5' splice site recognition. Furthermore, activation of a remote cryptic splice site at residue -29 of the upstream exon and missplicing of mRNA that is deleted for 29 nucleotides of the 5' exon are characteristic for this mutant. These results underscore the role of exon sequences in guiding the fidelity of the splicing reaction and they raise provocative questions about the alignment of introns within exon contexts that are consistent with accurate splicing and synthesis of an intact gene product.

[Key Words: Missplicing; 5' splice site; phage T4; group I intron; mutants; external guide sequence]

Received July 23, 1987; revised version accepted September 10, 1987.

The role of exon residues in the different RNA splicing reactions is becoming increasingly apparent. Exon sequences influence splice site selection in both spliceosome-dependent (Reed and Maniatis 1986) and self-splicing systems. The latter include the group I and group II introns (Michel and Dujon 1983), both of which have been shown to form pairings with their respective upstream exons (Perea and Jacq 1985; Been and Cech 1986; Waring et al. 1986; Jacquier and Rosbash 1986; Jacquier and Michel 1987). Interactions between internal guide sequences (IGSs) within the group I introns and upstream exon residues were predicted on the basis of sequence inspection and phylogenetic comparison (Waring et al. 1982; Michel et al. 1982). The instrumental role of these intron-exon pairings in guiding accurate 5' splice site cleavage for group I intervening sequences has since been proven experimentally by mutation and second-site reversion (Perea and Jacq 1985; Been and Cech 1986; Waring et al. 1986). More recently, the importance of intron-exon contacts has also been demonstrated for group II introns (Jacquier and Michel 1987).

The first step of the group I self-splicing reaction involves nucleophilic attack by a guanosine cofactor at the

5' splice site (Cech et al. 1981). It is this cleavage reaction that is guided by the IGS/upstream exon association (the P1 pairing) (Been and Cech 1986; Garriga et al. 1986; Inoue et al. 1986; Waring et al. 1986; Price et al. 1987). The 3'-OH of the upstream exon is then responsible for the simultaneous cleavage-ligation reaction at the 3' splice site, yielding the spliced exons (Cech et al. 1981). This reaction too is facilitated by interaction with the IGS, which appropriately positions the 3' end of the cleaved upstream exon for attack at the 3' splice site (Sullivan and Cech 1985; Been and Cech 1986; Waring et al. 1986).

The pivotal role of upstream exon sequences in determining the fidelity of 5' splice site selection is underscored by this genetic study of the group I phage T4 *td* intron. Among 13 nondirected splicing-defective mutations with impaired 5' splice site cleavage, only a single specificity mutant was found. This mutation was localized to the upstream exon whereas all the other mutations are found scattered throughout the 5' and 3' domains of the intron (Belfort et al. 1987). The mutation, a single base change 3 nucleotides from the 5' splice site, results in loss of 5' splice site recognition, activation of a cryptic splice site, and aberrantly spliced mRNA. The

missplicing event is responsible for deletion of 29 nucleotides from the exon ligation product. These results are discussed in regard to the exonic specificity determinants of splice site recognition. We refer to these critical exon residues as the external guide sequence (EGS). Furthermore, this work is viewed in the context of the appropriate alignment of introns within exon environments that are conducive to the precise splicing events required for orderly gene expression.

Results

*Isolation and preliminary characterization of splicing-defective *td* mutants*

The thymidylate synthase-encoding T4 *td* gene, which contains a 1016¹-nucleotide intron (Chu et al. 1984, 1986) was subjected to nondirected saturation mutagenesis. Thus, a library of 100 *td*⁻ mutants that are defective in thymidylate synthase (TS) production was isolated. This was achieved after hydroxylamine mutagenesis of pKK-TdΔ1-3, a plasmid containing a splicing-proficient *td* gene that had been deleted for 631

central intron residues. This plasmid contains intact exons, 166 5' intron residues, and 219 3' intron residues joined by an 8-nucleotide *Xba*I linker. The isolation of the *td*⁻ mutants was facilitated by selection for the TS⁻ phenotype after transformation of a *thyA*⁻ (TS⁻) host on trimethoprim-containing media in the presence of ampicillin and thymine (Belfort et al. 1987).

The initial screening for splicing-defective mutants was based on the *td* translation products. Expression of the wild-type *td* gene cloned into pKK-TdΔ1-3 results in the synthesis of three polypeptides. Translation of the pre-mRNA yields a 25-kD protein (NH₂TS) encoded by exon I, whereas the mature mRNA encodes the full-length 32-kD TS protein (Chu et al. 1985; Belfort et al. 1986). Both of these RNAs contain a downstream sequence that specifies an unrelated 11.4-kD peptide (X) (Belfort et al. 1986) (Fig. 1A,B). Band X is a useful feature for screening out transcription and RNA stability mutants, because unless transcripts extend beyond the 3' end of the *td* gene and persist in the cell, no product X is produced (for examples, see Fig. 1A). All of the polypeptides are easily distinguishable on a polyacrylamide gel, with an intense NH₂TS band (and band X) in conjunc-

¹The intron length was previously considered to be 1017 nucleotides. However, resequencing has revealed a single T at residue 93 of the intron rather than a doublet.

Figure 1. Identification of splicing-defective mutants. (A) NH₂TS assay. D1210*thyA* was transformed with pKK223-3 vector (O), pKK-TdΔ1-3 (WT), or TS⁻ mutants SC1–SC10 (from right to left). Transformants were treated with 2 mM IPTG for 3.5 hr to induce transcription from the *ptac* promoter. Coomassie Blue staining of a 15% acrylamide gel demonstrated two splicing-defective mutants, SC2 and SC4 (arrows). The positions of the three induced protein bands are marked on the right. A molecular weight scale in kilodaltons appears on the left. (B) Products of the *td* gene. Stippled boxes represent exon sequences, which are separated by the intron (heavy line). The 3' exon in this construct encodes a portion of another gene (X). Translation products that accumulate when splicing is defective are above the diagonal lines (NH₂TS and X), whereas products of the mature mRNA appear below (TS and X). Those mutants lacking band X are transcription or RNA stability mutants.

Figure 2. The oligodeoxynucleotide hybridization assay confirms splicing defects. RNA was isolated after 1 hr of transcriptional induction with 2 mM IPTG. D1210*thyA* transformants harbored pKK223-3 (O), pKK-TdΔ1-3 (WT), or putative splicing-defective derivatives of pKK-TdΔ1-3 (SC2, SC4, SC24, SC27, SC34, SC36, SC44, SC46, SC72, SC77, SC79, SC89, and SC99), as indicated by the corresponding designations on the left panel. RNA was immobilized and probed with ³²P-end-labeled oligonucleotides I, II, or III, shown below. Stippled boxes represent exons separated by the intron. The numbered bars represent the oligonucleotide probes. Probe I is a 22-mer complementary to the 5' intron 49 residues from the 5' splice site, probe II is a 25-mer complementary to the intron–exon II junction of the pre-mRNA, and probe III is a 24-mer complementary to the splice junction in the mRNA (Hall et al. 1987).

tion with the absence of the TS band acting as indicators of splicing defectiveness (Fig. 1A). The range of different polypeptide phenotypes is exemplified by the first 10 *td*[−] mutants screened (Fig. 1A). The protein profile of those mutants marked by arrows conforms to that which is predicted for the splicing-defective mutants. Twenty-five of the 100 *td*[−] mutants had increased NH₂TS/TS ratios. Of these, the 13 with the highest ratios were selected for further study.

To confirm the splicing-defective nature of these mutants, their RNAs, extracted from cells expressing the *td* gene, were screened by the dot-blot hybridization assay (Fig. 2) (Hall et al. 1987). Oligodeoxynucleotides specific for the intron–exon II junction of pre-mRNA (II) and for the splice junction of mRNA (III) demonstrated a buildup of pre-mRNA and only trace amounts of mRNA in mutants compared to wild type. The splicing-defective nature of all 13 of the mutants was thereby corroborated. Some of the mutants seem more defective than others, accumulating more pre-mRNA (probe II) and producing less ligated exon product (probe III). Among those few mutants giving a relatively weaker signal with the intron–exon II oligomer (probe II) is SC99. This result indicates some intermediate degree of cleavage activity at the 3' splice site of this mutant, despite the relatively low level of accurately spliced exons (probe III). SC99 is described in more detail below because of its other atypical properties.

Screening of mutants for cleavage at the 5' splice site

To examine the first step of splicing, RNA from cells transformed with each of the 13 *td* mutant plasmids was

used as template for cDNA synthesis with a primer that anneals to the intron 49 nucleotides from the 5' splice site (Fig. 3). For wild type, cDNAs correspond to linear intervening sequence, with stops in all of the lanes at the 5' end of the intron, and to the circular species, where the cDNA continues into the 3' end of the intron (Ehrenman et al. 1986). In contrast, for the mutants, exemplified by SC79 (Fig. 3A), the sequence reads from the 5' end of the intron into exon I of the pre-mRNA, indicating impaired cleavage at the 5' splice site. This was the case for all 13 mutants (Fig. 3B), where reduced cleavage at the 5' splice site is apparent from the diminished intensity of the cDNA banding in this region. Furthermore, the pattern for all of the mutants is similar to that of SC79, with one exception, SC99. For this latter case, a band of equivalent intensity to that at the wild-type 5' splice site was displaced by ~30 nucleotides up the gel (Fig. 3B, arrow). This result suggested unusual splice products for SC99. These data were corroborated by generating ddNTP sequence ladders, where all the mutants except SC99 behaved in identical fashion to SC79 (data not shown). Furthermore, the defect at the 5' splice site was confirmed for all the mutants by [³²P]GTP labeling reactions performed under self-splicing conditions (Price et al. 1987). Labeling of the 394 nucleotide linear intron band was reduced in intensity relative to wild type for each of the mutants. Although for SC99 the major labeled band was as intense as the wild-type linear intron band, it was larger by ~30 nucleotides (data not shown). The atypical splice products in this and the primer extension experiment (Fig. 3B) are both consistent with selection of a new 5' splice site for SC99.

Figure 3. Screening of mutants for splicing events at the 5' splice site. (A) Dideoxy sequence analysis. Total RNA from D1210*thyA* transformed with pKK-TdΔ1-3 (WT) or with splicing-defective mutant SC99 (MUT) was isolated 1 hr after induction with 2 mM IPTG. Primer extension sequencing reactions were performed with primer I (see Fig. 2 and Materials and methods) in the absence (0) or presence of the ddNTPs (A, C, G, and U correspond to the complement of the ddNTP in the reaction). Reading upwards from the bottom, the sequences diverge at the 5' end of the intron, at which position the wild type (but not the mutant) has multiple reverse transcriptase stops in all lanes. For wild type, the sequence beyond this point (horizontal line) corresponds to the 3' end of the circular form of the intervening sequence, whereas for the mutant, the sequence reads from the intron into exon I of the pre-mRNA. Those nucleotides that are obscured on the gel are bracketed. The predicted primer extension products for pre-mRNA (PRE), linear intron (L-IN), and circular (C-IN) are diagramed below, with the primer represented by a solid rectangle and the extension products by a wavy arrow. (B) Primer extension screening of mutant RNAs. RNA was prepared as above from splicing-defective mutants, SC2, SC4, SC24, SC27, SC34, SC36, SC44, SC46, SC72, SC77, SC79, SC89, and SC99 as well as from pKK-TdΔ1-3 (WT). cDNA synthesis was primed with the same oligonucleotide as in A, in the absence of ddNTPs. The primer extension of SC27 was repeated and was shown to have the same banding pattern as the other intron mutants, with greatly reduced band intensities at the 5' splice site relative to wild type. The only mutant that exhibited an atypical banding pattern was SC99 (arrow at position of major spurious band).

Sequence changes of the mutants

As a first step toward defining the residue changes for each of the mutants, the mutations were mapped by marker rescue. This was achieved by infecting cells harboring mutant plasmids with M13 hybrids containing *td* subfragments and by subsequently selecting for TS⁺ recombinants on media lacking thymine (Belfort et al. 1987). After localizing the mutations to deletion intervals about 150 nucleotides in length, these small regions were sequenced with *td*-specific primers, and the base change for each mutant was established. The mutations are shown in a linear representation in Figure 4A. Although all 13 mutants are defective at the earliest step in processing, cleavage at the 5' splice site, they lie scat-

tered throughout the 5' and 3' functional intron domains that flank the intron open reading frame (ORF) (Hall et al. 1987; Belfort et al. 1987). Again there is a single exception, SC99, which is an exon rather than an intron mutation. The lesion is a G → A change, three residues upstream of the 5' splice site.

Figure 4B shows the location of the mutations in relationship to the secondary structure model of the *td* intron. The distribution of the intron mutations in the context of the model is discussed by Belfort et al. (1987). We shall focus here on the single exon mutation, SC99, which is a critical residue of the P1 IGS-EGS pairing and which results in the accumulation of spurious splice products.

Figure 4. Distribution of *td* point mutations. (A) Linear map of mutations. Stippled boxes represent the exons and open boxes represent the intron ORF. Numbering commences at the first coded intron residue and is based on the undeleted wild-type *td* gene. (V) Location of the 631-nucleotide deletion within the ORF. (B) Intron mutations on the secondary structure map. The secondary structure, based on the model proposed by F. Michel (Shub et al., in press), is numbered like the linear map. The display of the intron and numbering of conserved structural elements, P1–P9.2, conforms to recently revised structural conventions (Burke et al. 1987). Arrows denote the position of the 5' and 3' splice sites. The intron sequence appears in upper-case letters, whereas exon residues are represented by lower-case letters. Letters indicating the conserved sequence elements P, Q, R, and S (Davies et al. 1982) are boxed. The start and stop codons of the ORF at residues 132 and 867, respectively, are shaded. The mutations, labeled as in A, are all G → A or C → U transitions.

Activation of a cryptic splice site and missplicing in mutant SC99

For SC99 the cDNA banding pattern was distinctive among the mutants (Fig. 3B). An intense reverse transcriptase stop, corresponding presumably to a 5' end, was observed approximately 30 nucleotides upstream of the normal 5' splice site. One explanation for this stop in SC99 might be utilization of a hidden or cryptic 5' splice site that lies upstream of the wild-type site. This notion is supported by the primer extension experiment shown in Figure 5, which also reveals the mutational change for SC99. The cDNAs were primed with the same oligonucleotide as used in Figure 3. Reading from the bottom of the autoradiogram in a 5' direction, wild-type and mutant sequences diverge at the 5' splice site. At this point there is a series of intense reverse transcriptase stops for wild type, after which the sequence reads into the 3' end of the intron. In contrast, for SC99

there is no cleavage activity detectable at the normal 5' splice site. Rather, the sequence continues into exon I for 29 nucleotides, at which point it is followed by strong reverse transcriptase stops characteristic of the 5' end of an excised intervening sequence.

If there is indeed a cryptic 5' splice site that yields a truncated upstream exon, what might the events be at the 3' splice site? While the dot-blot hybridization data (Fig. 2) indicate a moderate degree of cleavage activity at this site, can exon ligation occur? If so, does it occur at the normal 3' splice site? In addition to verifying the precise 5' splice site, primer extension analysis with an exon II primer answers all of the above questions (Fig. 6). The analysis was performed on wild type and two mutant RNAs (SC34 and SC99). Whereas all three RNAs revealed the same exon II sequence up to the intron junction, there was a dramatic divergence from this point. For SC34, which simply accumulates pre-mRNA, one reads from exon II into the intron. However, for wild

Figure 5. The 5' splice site of SC99. RNA from the missplicing mutant SC99 and from pKK-Td Δ 1-3 (WT) was prepared and analyzed by primer extension sequencing as for Fig. 3. The dashed arrow for SC99 represents the position of the 5' splice (5' SS) of the wild type. The upper arrow for SC99 marks the cryptic 5' splice site, 29 residues upstream of that in the wild type. The sequence of the excised intron–exon I hybrid molecule is shown on the right, with the noncoded G at its 5' end indicated in parentheses. (*) Mutated residue, a G $\rightarrow$ A change (compare with WT sequence in Fig. 6).

type, the sequence reads into the 3' end of exon I, yielding the normal mRNA junction. In sharp contrast, for SC99 the sequence skips the last 29 nucleotides of exon I, with the junction of the first residue of exon II (C) to the U residue upstream of the cryptic splice site in exon I. Thus, the truncated upstream exon effects cleavage and ligation at the normal 3' splice site, generating a misspliced exon product that is deleted for the terminal 29 nucleotides of the 5' exon. Sequences that are likely to be responsible for cryptic splice site activation and for the anomalous ligation reaction are considered below.

Discussion

Isolation of a splicing-defective specificity mutant

The experiments presented in this study emphasize the importance of upstream exon sequences in specifying the particular phosphodiester bond that constitutes the 5' splice sites and that is therefore destined for the first cleavage in the splicing pathway and subsequent ligation. Out of 100 *td*[−] mutants screened using the NH₂TS protein assay to indicate pre-mRNA buildup, 13 mutants were selected and demonstrated to have strong splicing-defective phenotypes (Figs. 1 and 2). Although all 13 mutants have impaired cleavage at the 5' splice site, only a single mutant, SC99, was shown to have undergone inaccurate 5' splice site cleavage (Fig. 3). By genetic mapping and DNA sequence analysis SC99 was also shown to be the only mutation located in the upstream exon (a G $\rightarrow$ A change at −3). The other mutations lie scattered throughout the 5' and 3' intron do-

main that have been implicated as the critical functional splicing components of the *td* intervening sequence (Fig. 4; Hall et al. 1987; Belfort et al. 1987). Primer extension analysis of SC99 RNA demonstrated activation of a distant cryptic splice site in the 5' exon, 29 residues upstream of the authentic 5' splice site (Fig. 5). Furthermore, the resulting truncated upstream exon was shown to be capable of accurate 3' splice site selection, yielding a ligation product deleted for the 29 nucleotides of the first exon (Fig. 6).

Knowledge of the genetic and functional lesions of SC99 provides insights into how this mutant passed the NH₂TS screening assay, despite the fact that bona fide NH₂TS cannot accumulate in this mutant. One must consider both translation from the unspliced pre-mRNA and translation from the misspliced exons. First, the G $\rightarrow$ A mutation at −3 results in a Gly $\rightarrow$ Ser change in the last residue of NH₂TS. This deduced change does not alter the number of residues (183 amino acids) in the molecule encoded by the unspliced precursor. Second, the missplicing event results in an exon ligation product predicted to encode a protein of 176 amino acids. This molecule would have the 10 carboxy-terminal residues of NH₂TS replaced by three amino acids contributed by downstream exon sequences, which thereafter specify a stop codon. This 176-amino-acid fusion molecule apparently has migration properties in a 15% polyacrylamide gel identical to those of authentic NH₂TS (data not shown). These two factors jointly contribute to mimicking high NH₂TS levels. In addition, the absence of a TS band and the presence of product X combine to satisfy all the criteria of a splicing-defective mutant.

Figure 6. Primer extension at the 3' splice site and splice junction of SC99. RNA was extracted from D1210*thyA* transformed with pKK-TdΔ1-3 (WT) or mutant derivatives SC34 or SC99. An exon II primer that anneals 53 nucleotides downstream of the 3' splice site was employed for primer extension dideoxy sequence analysis. Diagrams below the gels depict the form of the *td* transcript represented by the sequence. The solid rectangles represent the primer annealed to exon II. cDNAs are indicated by wavy lines with arrowheads pointing in the direction of reverse transcription. The shaded boxes depict exon sequences whereas solid lines represent intron sequence. The daggers mark the position of the cryptic splice site in precursor message (pre-mRNA) and in accurately processed message (mRNA). Dashed connecting lines in mRNA and misspliced RNA (ms-RNA) denote the joining of exon sequences. The junction sequences between exon I–exon II (EI–EII), intron–exon II (I–EII), and truncated exon I–exon II (EIIΔ–EII) are indicated to the left of the WT, SC34, and SC99 sequences, respectively.

The role of exon sequences in the specificity of splicing

Sequence inspection around the cryptic 5' splice site revealed a significant complementarity with the IGS. The putative pairing (cP1) that activates this cryptic site for the initial cleavage in the splicing reaction is shown in Figure 7. Strikingly, the cryptic splice site occurs downstream of three exon residues, GGU, that are identical to those in the normal P1 pairing. These residues are presumed to pair with the critical GCC residues in the IGS. Although the helical context of the GGU–GCC interaction for the normal and cryptic 5' splice sites is different (except for the additional C=G and U–A pairings indicated), the hypothetical pairing at the cryptic site is energetically equivalent to the P1 pairing (legend to Fig. 7). Furthermore, cleavage in cP1 is predicted to occur after the canonical U–G interaction (Michel et al. 1982; Waring and Davies 1984). It will be of interest to determine which of the cP1 base pairings are required for efficient activity at the cryptic 5' splice site.

The overall scheme for cryptic 5' splice site activation, accurate 3' splice site selection, and exon ligation is depicted in Figure 8. Thus, the intron is excised with 29 exon I residues at its 5' end, whereas the ligated exons are deleted for these corresponding residues. Although precise rate estimates are not yet available, the efficiency of the cleavage–ligation reaction for SC99

pre-mRNA is considerable. This is evidenced by primer extension analysis, which reveals efficient cleavage at the cryptic 5' splice site of SC99 (Figs. 3 and 5) and high levels of exon ligation (Fig. 6). However, cleavage at the 3' splice site does not seem to go to completion in SC99, as inferred from the intermediate hybridization signal with an intron–exon II probe (Fig. 2). This observation can be understood by comparing the IGS–EGS pairings for wild type and SC99 (Fig. 7). In the wild-type situation, the IGS–EGS pairing can be largely maintained after 5' cleavage (which occurs near the apex of the pairing), holding the free 5' exon stably in position to target the 3' splice site. In contrast, for SC99, this exon–IGS association is weakened after 5' cleavage since the cut site is within the mutant P1 helix. The free truncated upstream exon would therefore be in less stable association with the IGS, resulting in reduced 3' splice site cleavage.

The cryptic splice site is also activated, albeit at much lower levels, in wild-type transcripts synthesized in vitro. This is true for *td*Δ1-3 pre-mRNA as well as for the RNA precursor encoded by the full-length *td* intron (Belfort et al. 1987). Since the ΔGs of the wild-type P1 and cP1 are similar (legend to Fig. 7), a possible explanation for this observation is that the IGS pairs preferentially with the most proximal EGS. This may occur by a scanning mechanism whereby the intact wild-type EGS

sequence, cP1 has a ΔG of -4 kcal/mole (not shown). Using the revised parameters of Turner et al. (1987), these three values are -4.8 , -3.9 , and -3.3 kcal/mole, respectively. The ΔG of P1 for the SC99 sequence takes on a positive value by both methods of calculation.

is preferentially chosen, leading to cleavage predominantly at the wild-type splice site. An alternative explanation is that the normal P1 pairing and the IGS–cryptic EGS pairing are in equilibrium, but that the wild-type P1 helix is more efficient in directing the cleavage.

Remarkably, only one of the 13 mutants generated by nondirected mutagenesis exhibited altered splice site recognition. Furthermore, in an independent collection of 17 *td*[−] splicing-defective mutations generated in the

phage, again only a single 5' splice site specificity mutant was detected. Significantly, this was also an exon mutant, *tdH6*, with the base change at residue -2 , the nucleotide adjacent to that mutated in SC99 (Hall et al. 1987). Mutation *tdH6* activates the same cryptic splice site as SC99 (K. Ehrenman, D.H. Hall, and M. Belfort, in prep.). For the T4 *td*[−] mutant collection, mutations within the intron, including one within the IGS (N57; Hall et al. 1987), similarly reduced activity at the 5'

Figure 8. Overall scheme for normal and mutant splicing pathways. Exons are shaded and the intron is represented by a line. (A) Normal pathway. The IGS–EGS helix (O, P1 pairing) guides G attack and cleavage at the 5' splice site. Thereafter, the free upstream exon attacks the 3' splice site, resulting in ligated exons and excised intron. (B) Mutation. Mutation (X) of the EGS disrupts the normal P1 pairing, abolishing activity at the 5' splice site. (C) Mis-splicing. Pairing of the IGS to a secondary EGS (●) activates the cryptic splice site resulting in G attack at this site. The truncated upstream exon attacks at the normal 3' splice site, resulting in a ligated exon product that is deleted for the 3' end of the upstream exon (Δ). This deleted exon sequence is at the 5' end of the excised intervening sequence.

splice site, but none seemed to affect the accuracy of the residual reaction at this site. These observations reinforce the notion that the intron core structure constitutes the catalytic apparatus (Szostak 1986; Cech 1987), whereas exon residues that can form productive pairings with the IGS impart specificity to the reaction (Inoue et al. 1985; Been and Cech 1986; Garriga et al. 1986; Waring et al. 1986; Price et al. 1987).

Thus, in conclusion, mutation of the exon residues that normally pair with the IGS is capable of activating a cryptic splice site remote from the authentic 5' splice site. As a result, the attacking group is targeted to sequences within the upstream exon. Moreover, the truncated 5' exon has the appropriate specificity determinants that allow it to recognize and cleave the normal 3' splice site and effect exon ligation. This reaction is presumably achieved through the interaction with the IGS, as diagramed in Figures 7 and 8. These reactions appear analogous to those of a 5' exon mutation isolated by Been and Cech (1986) in the *Tetrahymena thermophila* large rRNA gene. In this mutant, the 5' splice site slips by two residues into the upstream exon, resulting in deletion of a dinucleotide from the exon ligation product. Additionally, Price et al. (1987) have recently shown loss of splicing specificity by removal of all but two of the natural 5' exon residues flanking this intron. In this case 5' splice site activation in the 3' exon was demonstrated, underscoring the role of exon sequences in guiding splice site selection.

Intron-exon alignment, intron mobility, and productive gene expression

The impact of the exon environment on the outcome of a splicing event presents a paradox in terms of the implied mobility of group I and group II introns (Dujon et al. 1986; Jacquier and Michel 1987). Clearly, if a single base change in the upstream exon can result in a 29-nucleotide deletion in the mature mRNA, ectopic positioning of introns during horizontal transfer would have catastrophic consequences in terms of gene expression. One of two scenarios may be envisaged to resolve this paradox: (1) Functional introns may be confined to integrating within exon sequences that can support specific splicing events that yield functional gene products, or (2) introns may carry with them an exon environment that is consistent with splicing. In those cases in which the new exon sequences are compatible with gene function, a nondisruptive transposition event into the new gene will be achieved.

In this regard, it is interesting to note that the variable occurrence of the T-even phage introns strongly suggests their evolutionary instability and possible mobility (Pedersen-Lane and Belfort 1987; unpubl.). Furthermore, the immediate exon contexts of the three homologous T4 introns of the *td*, *nrdB*, and *sunY* genes are extremely similar, despite the different functions of their gene products (Shub et al. 1987). Whether these exon sequence identities are the cause or the effect of the localization of the introns in the three genes remains a matter of speculation.

Materials and methods

Plasmids, bacterial strains, and media

The "wild-type" *td* plasmid used in this study, pKK-TdΔ1-3, contains a 2-kb *EcoRI*-*HpaI* *td* fragment (Belfort et al. 1987) under transcriptional control of the *ptac* promoter in vector pKK223-3 (Brosius and Holy 1984). The fragment of the *td* gene in pKK-TdΔ1-3 contains intact exons but has a 631-nucleotide deletion in the intron ORF (Belfort et al. 1987). Mutant derivatives of the *td* gene were isolated from this construct, which is fully splicing proficient. These *td* mutants (designated SC) were isolated on selective media after hydroxylamine mutagenesis of pKK-TdΔ1-3 and transformation of D1210*thyA* (Belfort et al. 1987). *Escherichia coli* strain D1210*thyA* was prepared in this laboratory from D1210, a *lacI^q* derivative of HB101 (Brosius and Holy 1984). Growth medium used for the protein and RNA analyses was TBYESA, which is tryptone broth (1.0% Bacto-tryptone and 0.5% NaCl) supplemented with 0.5% yeast extract, 50 μg/ml thymine, and 100 μg/ml ampicillin.

NH₂TS assay

Transformants were grown in TBYESA until mid-log phase at 37°C. Isopropylthio-β-galactoside (IPTG) was then added to a concentration of 2 mM. Cultures were induced for 3.5 hr, and the cells pelleted and then washed with TM (10 mM Tris-HCl, pH 7.4; 10 mM MgCl₂). Washed pellets were lysed in stop-load buffer and loaded onto a 15% bis-acrylamide gel (Laemmli 1970). The gel was stained with Coomassie Blue for 30 min at 60°C and then destained overnight in a solution containing 10% ethanol and 10% acetic acid.

Dot-blot RNA assay

Transformants were grown and induced with IPTG for 1 hr as described above. Total cellular RNA was extracted by the lysozyme freeze-thaw method followed by phenol extraction and ethanol precipitation as previously described (Belfort et al. 1985). Hybond-N filters (Amersham) were cut to size and spotted with 800 ng of total cellular RNA from induced cells. The filters were UV irradiated for 5 min to fix the RNA. End-labeling of oligodeoxynucleotides using [γ -³²P]ATP and hybridization were carried out according to Hall et al. (1987). Filters were washed three times for 30 min each in 6× SSC at 23°C before autoradiography.

Primer extension analysis

Total cellular RNA was extracted from transformants as described above. Synthetic oligodeoxynucleotide primers were end-labeled with T4 polynucleotide kinase in the presence of [γ -³²P]ATP. The primers were either a 20-mer complementary to exon II, 53 nucleotides downstream for the 3' splice site, or a 22-mer complementary to the intron, 49 nucleotides downstream from the 5' splice site (see below). Annealing of 0.2 pmole of RNA with 0.6 pmole of labeled oligonucleotide was accomplished by heating to 60°C for 3 min, then quick-freezing. Extension with dNTPs and AMV reverse transcriptase (Life Sciences) was performed at 48°C. ddNTPs were added to the reactions as indicated. The cDNA products were separated on 8% polyacrylamide-urea gels.

Oligodeoxynucleotide probes and primers

The oligonucleotide probes I, II, and III used for hybridization analysis correspond to oligomers 1, 3, and 4, respectively, of

Hall et al. (1987). cDNA synthesis was primed from intron or exon II sequences using synthetic oligomers 1 and 7, respectively, as detailed by Hall et al. (1987).

Acknowledgments

We greatly appreciate the help of Joan Pedersen-Lane with several of the primer extension analyses. We thank Jill Galloway Salvo and Peter DiMaria for their helpful comments on the manuscript. This work was supported by grants from the National Science Foundation (DMB-8502961) and the National Institutes of Health (GM 33314).

References

- Been, M.D. and T.R. Cech. 1986. One binding site determines specificity of *Tetrahymena* pre-rRNA self-splicing, trans-splicing and RNA enzyme activity. *Cell* **47**: 207–216.
- Belfort, M., J. Pedersen-Lane, D. West, K. Ehrenman, G. Maley, F. Chu, and F. Maley. 1985. Processing of the intron-containing thymidylate synthase (*td*) gene of phage T4 is at the RNA level. *Cell* **41**: 375–382.
- Belfort, M., J. Pedersen-Lane, K. Ehrenman, F.K. Chu, G.F. Maley, F. Maley, D.S. McPheeters, and L. Gold. 1986. RNA splicing and *in vivo* expression of the intron-containing *td* gene of bacteriophage T4. *Gene* **41**: 93–102.
- Belfort, M., P.S. Chandry, and J. Pedersen-Lane. 1987. Genetic delineation of functional components of the group I intron in the phage T4 *td* gene. *Cold Spring Harbor Symp. Quant. Biol.* **52**: (in press).
- Brosius, J. and A. Holy. 1984. Regulation of ribosomal RNA promoters with a synthetic lac operator. *Proc. Natl. Acad. Sci.* **81**: 6929–6933.
- Burke, J.M., M. Belfort, T.R. Cech, R.W. Davies, R.J. Schweyen, D.A. Shub, J.W. Szostak, and H.W. Tabak. 1987. Structural conventions for group I introns. *Nucleic Acids Res.* **15**: 7217–7221.
- Cech, T.R. 1987. The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**: 1532–1539.
- Cech, T.R., A.J. Zaug, and P.J. Grabowski. 1981. *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**: 487–496.
- Chu, F.K., G.F. Maley, F. Maley, and M. Belfort. 1984. An intervening sequence in the thymidylate synthase gene of bacteriophage T4. *Proc. Natl. Acad. Sci.* **81**: 3049–3053.
- Chu, F.K., G.F. Maley, M. Belfort, and F. Maley. 1985. *In vitro* expression of the intron-containing gene for T4 phage thymidylate synthase. *J. Biol. Chem.* **260**: 10680–10688.
- Chu, F.K., G.F. Maley, D.K. West, M. Belfort, and F. Maley. 1986. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. *Cell* **45**: 157–166.
- Davies, R.W., R.B. Waring, J.A. Ray, T.A. Brown, and C. Scazzocchio. 1982. Making ends meet: A model for RNA splicing in fungal mitochondria. *Nature* **300**: 719–724.
- Dujon, B., L. Colleaux, A. Jacquier, F. Michel, and C. Monteilhet. 1986. Mitochondrial introns as mobile genetic elements: The role of intron-encoded proteins. In *Extrachromosomal elements in lower eukaryotes* (ed. R.B. Wickner, A. Hinnebusch, A.M. Lambowitz, I.C. Gunsalus, and A. Hollaender), pp. 5–27. Plenum Press, New York.
- Ehrenman, K., J. Pedersen-Lane, D. West, R. Herman, F. Maley, and M. Belfort. 1986. Processing of phage T4 *td*-encoded RNA is analogous to the eukaryotic group I splicing pathway. *Proc. Natl. Acad. Sci.* **83**: 5875–5879.
- Garriga, G., A.M. Lambowitz, T. Inoue, and T. Cech. 1986. Mechanism of recognition of the 5' splice site in self-splicing group I introns. *Nature* **322**: 86–89.
- Hall, D.H., C.M. Povinelli, K. Ehrenman, J. Pedersen-Lane, F. Chu, and M. Belfort. 1987. Two domains for splicing in the intron of the phage T4 thymidylate synthase (*td*) gene established by non-directed mutagenesis. *Cell* **48**: 63–71.
- Inoue, T., F.X. Sullivan, and T.R. Cech. 1985. Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: Oligonucleotides can function as 5' exons. *Cell* **43**: 431–437.
- . 1986. New reactions of the ribosomal RNA precursor of *Tetrahymena* and the mechanism of self-splicing. *J. Mol. Biol.* **189**: 143–165.
- Jacquier, A. and F. Michel. 1987. Multiple exon-binding sites in class II self-splicing introns. *Cell* **50**: 17–29.
- Jacquier, A. and M. Rosbach. 1986. Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5' exon–intron interaction. *Science* **234**: 1099–1104.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.
- Michel, F. and B. Dujon. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial, chloroplast-, and nuclear-encoded members. *EMBO J.* **2**: 33–38.
- Michel, F., A. Jacquier, and B. Dujon. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**: 867–881.
- Pedersen-Lane, J. and M. Belfort. 1987. Variable occurrence of the *nrdB* gene in the T-even phages suggests intron mobility. *Science* **237**: 182–184.
- Perea, J. and C. Jacq. 1985. Role of the 5' hairpin structure in the splicing accuracy of the fourth intron of the yeast *cob-box* gene. *EMBO J.* **4**: 3281–3288.
- Price, J.V., J. Engberg, and T.R. Cech. 1987. 5' Exon requirement for self-splicing of the *Tetrahymena thermophila* pre-ribosomal RNA and identification of a cryptic 5' splice site in the 3' exon. *J. Mol. Biol.* **196**: 49–60.
- Reed, R. and T. Maniatis. 1986. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **46**: 681–690.
- Shub, D.A., J. Gott, M.Q. Xu, B.F. Lang, F. Michel, J. Tomaschewski, J. Pedersen-Lane, and M. Belfort. 1987. Structural conservation of three homologous introns of phage T4 and the group I introns of eukaryotes. *Proc. Natl. Acad. Sci.* (in press).
- Szostak, J.W. 1986. Enzymatic activity of the conserved core of a group I self-splicing intron. *Nature* **322**: 83–86.
- Turner, D.H., J. Sugimoto, J.A. Jaeger, C.E. Longfellow, S.M. Freier, and R. Kierzek. 1987. Improved parameters for prediction of RNA structure. *Cold Spring Harbor Symp. Quant. Biol.* **52**: (in press).
- Waring, R.B. and R.W. Davies. 1984. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing: A review. *Gene* **28**: 277–291.
- Waring, R.B., P. Towner, S.J. Minter, and R.W. Davies. 1986. Splice-site selection by a self-splicing RNA of *Tetrahymena*. *Nature* **321**: 133–139.
- Zuker, M. and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**: 133–148.

Activation of a cryptic 5' splice site in the upstream exon of the phage T4 td transcript: exon context, missplicing, and mRNA deletion in a fidelity mutant.

P S Chandry and M Belfort

Genes Dev. 1987, 1:

Access the most recent version at doi:[10.1101/gad.1.9.1028](https://doi.org/10.1101/gad.1.9.1028)

References

This article cites 29 articles, 7 of which can be accessed free at:
<http://genesdev.cshlp.org/content/1/9/1028.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).