Mechanism of 3' splice site selection by the catalytic core of the sunY intron of bacteriophage T4: the role of a novel base-pairing interaction in group I introns

François Michel,1 Pierre Netter,1 Ming-Qun Xu,2 and David A. Shub2

1Centre de Génétique Moléculaire du Centre National de la Recherche Scientifique, Laboratoire Associé à l'Université Pierre et Marie Curie, 91190 Gif-sur-Yvette, France; 2Department of Biological Sciences and Center for Molecular Genetics, State University of New York at Albany, Albany, New York 12222 USA

The catalytic core of the sunY intron of bacteriophage T4 is separated from its 3' exon by 837 nucleotides, most of which are part of an open reading frame [ORF]. Here, we report that transcripts truncated within the sunY ORF self-splice in vitro to a variety of sites in the segment immediately 3' of the core. Recognition of these proximal splice sites is shown to depend on (1) the presence on the intron side of a terminal G, which must not be part of a secondary structure; and (2) the ability of the penultimate intron nucleotide to base-pair with a 3' splice site-binding sequence [3'SSBS] located within the core. The counterpart of the 3'SSBS can be identified in most group I introns. The possible significance of such alternative splicing events for in vivo expression of intron-encoded proteins is discussed.

[Key Words: Group I intron; RNA splicing; splice site-binding sequence]

Received January 16, 1990; revised version accepted February 27, 1990.

Group I introns are found in genes of eukaryotic organelles, the nuclear rRNA genes of some protists, as well as certain bacteriophage [for review, see Cech 1988]. Because a number of these introns can self-splice in vitro, the specificity that determines their cleavage-ligation sites must arise from interactions within the precursor RNA. A representation of the established secondary structure elements of group I introns can be seen in Figure 1, which depicts the sunY intron of bacteriophage T4.

There is a simple rule that determines the 5' splice site of all group I introns: The last nucleotides of the 5' exon are involved in base-pairing [called P1], with a complementary sequence located within the intron, 5' to its active core. In almost all cases [including sunY], the exon ends with a U, which is paired to a G in the intron. This rule, which was originally proposed on the basis of sequence comparisons [Davies et al. 1982; Michel et al. 1982], has subsequently been experimentally confirmed [Been and Cech 1986; Waring et al. 1986]. The P1 pairing may be sufficient to specify the 5' cleavage site, because point mutations in the exon [that destabilize P1] can activate cryptic 5' splice sites that mimic the original pairing [Chandry and Belfort 1987; Price et al. 1987].

The specificity at the 3' splice site is not as well understood. All group I introns end with a G, and changing that nucleotide to U or C in the intron of the large rRNA of Tetrahymena thermophila allows only residual splicing [Price and Cech 1988]. Also, the 2 nucleotides immediately preceding the 3'-terminal G of the Tetrahymena intron have been shown to contribute significantly to binding of the 3' splice junction by the catalytic core [Tanner and Cech 1987]. Indeed, mutating the penultimate residues of the intron results in slowing the rate of splicing, without changing its specificity [Price and Cech 1988]. In addition, Davies et al. [1982] noted that 5' to the intron residues that pair with the 5' exon to form helix P1, there are usually several additional residues that can pair [P10] with the first several residues of the 3' exon. Most group I introns would thus include an internal guide sequence [IGS], which, by pairing with both 5'- and 3'-exon residues [P1 and P10], would align the two exons for the ligation step [see Fig. 2a]. However, the P10 pairing of the Tetrahymena intron can be disrupted by mutations, some having little effect on the rate of splicing and none on the choice of the 3' splice site [Been and Cech 1985].

That the P10 interaction alone is insufficient to specify the 3' splice site is perhaps most apparent in the sunY intron of bacteriophage T4, where >800 nucleotides [most of them comprising an open reading frame [ORF]] separate the catalytically active core structure from the 3' splice site [Shub et al. 1988; Doudna and Szostak 1989; Xu and Shub 1989]. Figure 2a shows a representation of the sunY precursor RNA prior to 3' cleavage and exon ligation. The pairing between the IGS
and exon sequences is interrupted by an A-A mismatch between P1 and P10, and the latter pairing comprises only 3 bp. There are, moreover, several sequences within [or immediately preceding] the sunY ORF that could substitute for the 3' strand of the putative P10 pairing. The closest of these to the core occur after G_{235} [the residue preceding the start of the ORF], where the next 5 residues [AUGAA] are identical to those at the beginning of exon 2, and after G_{227} [within the region of complementarity to 16S RNA (ribosome-binding site, RBS) preceding the ORF], where a G-A mismatch would be followed by an extended P10 pairing [Fig. 2b]. Although inclusion in a stable secondary structure [see Fig. 2c] may prevent their utilization in full-length transcripts, these potential splice sites may be transiently exposed during transcription of the ORF [see Fig. 2b]. Yet splicing occurs both in vitro and (as far as we can tell) in vivo at a unique site 3' of the ORF.

We have sought to obtain clues to the rules governing 3' splice site selection by studying reactions of transcripts containing a complete sunY catalytic core but truncated within the ORF. Thus, the ability of sequences [that are transiently exposed during transcription] to function as 3' splice sites can be examined in the absence of competition from the authentic, distal, site. We have been able to show that [1] a variety of proximal splicing events can be detected in vitro, both in wild-type molecules and molecules containing point mutations; [2] a base-pairing, previously undescribed for group I introns, between the penultimate nucleotide \( 5' \)

RTX END
Group I intron 3' splice site selection

Figure 2. (See following page for legend.)
of the terminal G) and a residue in the core, is a major determinant of the 3′ splice site.

Results

**Truncated transcripts of the sunY intron yield novel splicing products**

In vitro self-splicing experiments are usually initiated by transferring purified, unreacted full-length transcripts from nonpermissive to permissive reaction conditions. This is a reasonable way to proceed as long as the 3′ intron–exon junction lies immediately downstream of the catalytic core, as is the case for most group I introns. However, 837 nucleotides separate the catalytic core of intron sunY (defined as ending with stem P9, its last evolutionarily conserved element; see Figs. 1 and 2a) from its 3′ extremity. This leaves considerable time during in vivo synthesis for incomplete transcripts to participate in minor reactions, at sites that may be inaccessible in full-length transcripts having reached confirmational equilibrium. Therefore, only by assessing the reactivity of truncated transcripts of the sunY intron can a full description of its splicing potential be obtained. To conveniently generate truncated transcripts, we subcloned the 526-bp AccI–SspI restriction fragment from sunY (containing a short 5′ exon sequence and 500 intron nucleotides) into plasmid expression vectors (see Fig. 3).

Transcripts ending at the DraI cleavage site, 135 nucleotides 3′ of stem P9, when incubated under splicing conditions, were found to yield several novel products of ∼140, 145, 240, and 245 nucleotides [Fig. 4]. These products were identified by gel extraction, followed by sequencing with reverse transcriptase. The bands at ∼140 and 145 nucleotides correspond to ligated exon products resulting from cleavage on the 3′ side of the G residues at positions 243 and 238 from the beginning of the intron [Fig. 2c], respectively, followed by ligation to the 5′ exon [Fig. 5A,B]. The bands at ∼245 and 240 nucleotides presumably result from the corresponding excised linear introns, as their 5′ extremities coincide with the 5′ exon–intron junction of sunY (not shown). Transcripts terminating at the SspI site (see Fig. 1) yielded essentially similar results [data not shown].

Production of the three or four molecular species is under GTP control through the intermediate of 5′-exon cleavage: Prolonged incubation in the absence of GTP results in only traces of free 5′ exon and no detectable ligation products [data not shown]. It was also determined that the kinetics of cleavage at the bonds following G243 and G238 parallel those of 5′-exon release: The ratio of ligated exons to free 5′ exon remains constant with time [data not shown].

However, production of ligated exon and excised intron products is not stoichiometric with that of 5′ exon. Whereas conversion of precursor RNA to intron/3′ exon is 85% complete in 5 min at 45°C, the fraction of 5′ exon that has become ligated does not exceed ~0.25 and does not increase on prolonged incubation. Rather, prolonged incubation results in the appearance of multiple [at least five] circular intron products [Fig. 4]. Two of them were extracted and reverse-transcribed [data not shown]. In one case, the G238 residue was found to have become covalently linked with U9, whereas in the other one, G243 was joined to U7. Because both U residues must have been preceded by a G (G23 and the G added to the 5′ extremity of the intron in the first step of the reaction, respectively), both circularization events presumably reflect the “G-exchange” reaction known to be catalyzed by at least some group I introns [Inoue et al. 1986; Van der Horst and Tabak 1987].

Finally, two additional products of ∼88 and 94 nucleotides were visible on gels after incubation at pH 8.2 [Fig. 4] but not at pH 7.0 [data not shown]. They were shown by reverse transcription to correspond to free 3′ exons, liberated by cleavage of the bonds following G243 and G238.

Alternate P9.1 pairings govern the choice of intron–3′ exon junctions by the core of the sunY intron

As shown in Figures 1 and 2a, truncation of sunY transcripts within ORF 55.12 results in the loss of the 3′ components of the composite P9.1 structure. However, the nucleotide sequence lying immediately 3′ of the P9 stem–loop structure can readily be refolded into two different, mutually exclusive P9.1a-like double-stranded segments (Fig. 2b and c). Both of these pairings include the 5′ branch of the long-distance P9.1a pairing in Figure 3.

---

**Figure 2.** Potential for secondary structure formation by sequences lying downstream of the core of the sunY intron. All transcripts are shown after having undergone cleavage at the 5′ intron/exon junction. Potential interactions between the 5′ and 3′ splice sites are illustrated. (a) Full-length transcript. Lowercase letters: exon sequences. Double-stranded segments are labeled according to conventions of Burke et al. (1987). X[1–2] and Y[1–3] are potential components of a P9.1a helix [see b and c]. Boxed elements are conserved in all three T4 introns [see Shub et al. 1988]. The heavy arrow indicates the intron/3′-exon junction [distal splice site]. ORFs 55.13 and 55.11 [Tomaszewski and Rüger 1987] are the 5′ and 3′ exon sequences, respectively. Putative start and stop codons and a potential RBS for ORF 55.12 are shown. G residues that are followed by sequences similar to that at the distal splice site are indicated with an asterisk. The interaction of the penultimate intron nucleotides with the 3′ SSBS is shown as a dashed line. (b) Transcript ending within ORF 55.12, folded into state I [see text and d]. Arrows indicate observed 3′ splice sites [labeled according to intron terminal G], with preference indicated by width of arrow. Open arrows indicate sites that are used only in transcripts in which position 177 is mutated from A to G [the site following G238 is also used when position 234 is mutated to U, see Fig. 4 and text]. (c) Same as b, but folding is into state II. (d) Mutations introduced into the sunY intron. I and II refer to two possible, mutually exclusive states of the RNA sequence lying immediately 3′ to the core of the sunY intron [cf. with b and c]. Arrows indicate mutated sites with position [numbered from first intron nucleotide] and substituted nucleotide[s] indicated. GAA207 refers to substitution of three consecutive nucleotides (UGC), starting at position 207.
Group I intron 3' splice site selection

Figure 3. Schematic representation of template for in vitro transcription and of products of in vitro splicing reactions. (A) Plasmid pGEM-3/T4AS, formed by insertion of an Accl–SspI restriction fragment from the sunY locus into pGEM-3 (see Experimental procedures). Hatched regions are T4 sequences flanking the intron catalytic core (open areas). Solid areas are vector sequences. Numbers indicate distance from the first intron nucleotide. (B) Intermediate of in vitro splicing reactions. (C) Products of in vitro splicing reactions (see Figs. 4 and 6).

2a (sequence element X1) as part or all of their 5' component. They differ, however, in their 3' branch, which reads AUAAUA (sequence element Y1) in one case (Fig. 2b), and AAACG (sequence element Y2) in the other (Fig. 2c), versus AUAAACCG (element Y3) for the long-range P9.1a pairing of full-length transcripts (Fig. 2a). Although the X1–Y1 pairing should form first, as RNA synthesis proceeds, it should soon be replaced by the X1–Y2 pairing (state II in Fig. 2d), which is overwhelmingly favored at equilibrium (only the X1–Y2 pairing can be combined with the short-range P9.1b hairpin structure, which is highly stable, and is actually part of the P9.1 complex of full-length transcripts; see Fig. 2a).

Therefore, one could hardly gain access to the earliest, most proximal splicing events possibly associated with state I, unless mutations were introduced that stabilize the X1–Y1 pairing at the expense of the X1–Y2 and P9.1b structures. The necessary nucleotide replacements (Fig. 2d) were carefully chosen so that at the same time, they would weaken or destroy one or more base pairs of the component structural elements of state II and stabilize state I, either by extending or strengthening the X1–Y1 helical stem or, in the case of mutant combination 206U–210G, converting the CCUGGU-terminal loop to the thermodynamically highly favored CUUCGG sequence (see Tuerk et al. 1988).

A total of eight different truncated transcripts, carrying various combinations of the mutations shown in Figure 2d, were assayed for their ability to generate splicing products. Six of these are shown in Figure 4, and all of them are splicing-proficient, at levels comparable to those observed with the wild-type sequence. The extent of 5' cleavage is essentially the same for all transcripts, whereas the ratios of 3' cleavage to 5' cleavage do not differ markedly, whether between mutants or with respect to wild-type molecules. However, a variety of spliced exon and excised intron products are observed. In addition, all mutant transcripts generate multiple circular forms of the excised introns, as well as an increased variety of linear ones under prolonged incubation.

Transcripts carrying a single mutated nucleotide were not expected to differ from the wild type, because of insufficient effects on the relative stabilities of states I and II (Table 1), and this is indeed the case for mutants G210 (Fig. 4, no. 1) and C216 (data not shown). On the other hand, the most heavily mutated transcripts are seen to generate novel splicing patterns: The wild-type bands disappear and are replaced by several new ones, corre-
Michel et al.

Figure 4. Patterns of splicing products generated by wild-type and mutant transcripts of the sunY intron truncated within ORF 55.12. Autoradiograph depicts electrophoresis (on a 6% acrylamide/urea gel) of the products of a 45-min reaction at 37°C (pH 8.2). Gel-purified precursor RNAs generated by in vitro transcription of DraI-cleaved wild-type and mutant derivatives of pGEM-3/T4AS plasmid DNA were preincubated for 20 min in 50 mM Tris-HCl, 50 mM NH4Cl, 30 mM MgCl2, 10 mM spermidine, and 2 mM GTP at 37°C (final pH 8.2). The reaction was started by addition of GTP to a final concentration of 2 mM. Mutants are numbered as in Table 1 (two independent constructs of triple mutant 5 are shown). Sizes of products identified by reverse transcription of gel-extracted material are shown at right. Numbers at left identify splicing products formed at 3' splice sites following Gs with indicated coordinates. Minus-GTP control lanes (data not shown) showed only trace amounts of intron–3' exon and 5' exon and no other splicing products. Double bands of 3'-exon hydrolysis products and ligated exon species result from the tendency of phage SP6 RNA polymerase to add untemplated nucleotides to the 3' ends of transcripts.

Changing one nucleotide in the J7-9 core segment changes the consensus 3' splice sequence from UG to CG

It is clear from our experiments (Fig. 4) that 3' cleavage by truncated transcripts of the sunY intron occurs preferentially after the UG dinucleotide. Of the four major ligated exon products we characterized, three were generated by cleavage of a UGA sequence, whereas the fourth one corresponds to cleavage of UG227G (UGA is also the sequence at the distal 3'-splice site in full-length transcripts; see Shub et al. 1988). In contrast, the AG227G, GG227U, and GG227A sequences were used rather poorly, and CGA (at positions 224–226 and 234–236) was not used at all, unless the C had first been mutated to a U.

In spite of their involvement in 3' splice site recognition, the penultimate nucleotides of group I introns vary even within the same subgroup of introns (see Cech 1988). This makes it feasible to look for covariation with one or several core residues, which may be indicative of direct physical interactions. We thus observed that the first two or three residues of the J7-9 segment connecting core helices P7 and P9—henceforth designated as 3'SSBS (3' splice site-binding sequence)—seemed to be constrained by the nature of the two residues preceding the 3'-terminal G (Fig. 7). Even though these two groups of residues may not always form Watson–Crick pairings, there are nevertheless clear instances of coordinated evolution to maintain complementarity between
Group I intron 3' splice site selection

Figure 5. Identification of ligated exon products generated by wild-type and mutant sun\textit{Y} transcripts truncated within ORF 55.12. Sequence ladders were generated by reverse transcription of gel-extracted material (see Experimental procedures): \{A and B\} Wild-type splicing products; \{C and D\} splicing products of mutant 6 [207GAA-210G-234U]; \{E\} splicing products generated by wild-type precursor RNA. Lanes are labeled with the complement of the dideoxynucleotide used in the reaction. Horizontal arrows indicate splice junctions. Spliced exon products are designated by positions of the intron-encoded G residues immediately preceding their 3' splice junctions.

Table 1. Predicted relative stabilities of states I and II of the P9.1 hairpin structure in truncated transcripts

| Sites | wt | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|----|---|---|---|---|---|---|---|
| 206   | C  | C | C | C | C | C | C | C |
| 210   | U  | U | G | U | G | G | G | G |
| 216   | U  | C | U | U | U | C | C | U |
| 234   | C  | C | C | U | U | C | U | C |
| IIa   | -8.9 | -8.9 | -4.0 | -8.9 | -4.0 | -4.0 | -4.0 | +4.2 |
| IIb   | -2.8 | -2.8 | -2.8 | -0.9 | -0.9 | -2.8 | -0.9 | -0.9 |
| II    | -11.7 | -11.7 | -6.8 | -9.8 | -4.9 | -6.8 | -4.9 | +3.3 |
| I     | -0.9 | -3.4 | -1.6 | -0.9 | -1.6 | -4.1 | -4.1 | -1.6 |
| II−I  | -10.8 | -8.3 | -5.2 | -8.9 | -3.3 | -2.7 | -0.8 | +4.9 |

Sites that were changed in mutant constructs are underlined. IIa and IIb refer to terminal (nucleotides 205–228) and basal (nucleotides 198–204 and 229–235) components, respectively, of structure II (see Fig. 2d). Stabilities are ΔG values at 37°C, expressed in kcal/mole in 1 M NaCl and calculated according to Turner et al. (1988).
Figure 6. The pattern of splicing products of truncated transcripts is dependent on the interaction of a single core nucleotide with the penultimate intron nucleotide. Autoradiograph depicts electrophoresis (on a 6% acrylamide/urea gel) of the products of a 5-min reaction at 45°C (pH 7.0). Gel-purified precursor RNAs generated by in vitro transcription of DraI-digested pTZ19U derivatives of our initial pGEM3/T4AS sunY construct were preincubated for 10 min in 50 mM Tris-HCl (pH 7.0), 50 mM NH₄Cl, 30 mM MgCl₂, and 0.02% SDS. The reaction was started by addition of GTP at a final concentration of 2 mM. Mutants are numbered as in Table 1. On the right side, ligated exon products formed at 3’ splice sites following Gs with indicated coordinates are identified, together with the nucleotides preceding and following these Gs. Note that the two transcripts containing mutation combination 4 differed only by the presence of the mutation A177 to G. Double bands of ligated exon species result from the tendency of phage T7 RNA polymerase to add untemplated nucleotides to the 3’ ends of transcripts.

The 3’SSBS and the penultimate intron nucleotides, [e.g., among close relatives of the Tetrahymena ribosomal intron]. Particularly striking is the fact that although the first nucleotide in J7-9 is almost always a purine, it is C in the four introns that have a G as their penultimate nucleotide [cf. the tRNA Leu introns in the Cyanophora paradoxa (C.p. tLeu) cyanelle and the Marchantia polymorpha (M.p. tLeu) chloroplast with those in higher plant chloroplasts].

In most members of subgroup IA [to which the sunY intron belongs], the sequence complementary to the penultimate intron nucleotides seems to consist of residues 2 and 3 of the J7-9 segment [A177 and A178 in sunY; see Figs. 1 and 2a]. We therefore changed residue A177 to G in the hope of changing the specificity of recognition of 3’ splice sites by the core of the sunY intron. To gain access to the entire segment extending from positions 218 to 244 and beyond, the 177G mutation was combined with mutant 4 [210G–216C]. As shown in Figure 6, incubation of the corresponding truncated transcript under splicing conditions generates two new ligated exon products, the lengths of which show that they result from cleavage of the two CGA sequences at 224–226 and 234–236. The faster migrating of these new species [transcribed in the absence of isotopic label] was extracted from a gel and sequenced by reverse transcription [Fig. 8]. It is clearly the product of ligation of the 3’-terminal U of exon 1 to A236. The UGG and UGA sequences at 242–244 and 237–239 are still cleaved, but to a lesser extent.

There are two peculiarities of transcripts carrying the mutation of A177 to G. First, the coupling between 5’ and 3’ cleavage is somewhat better, with up to 55% of cleaved 5’ exon ending up as ligated products [data not shown]. In addition, the linear intron forms undergo additional rounds of 3’ cleavage: Prolonged incubation (for 1 hr or more, data not shown) results in a single intron band corresponding to a molecule of ~220 nucleotides, whereas the pattern of ligated exon bands remains unaltered. This is in marked contrast to the wild-type and all other mutants we examined, where an array of intron products are found whose relative abundances reflect those of the corresponding ligated exons.

Discussion

Selection of 3’ splice sites by group I introns: the role of the 3’SSBS

From our experiments with truncated transcripts of the sunY intron, three conditions emerge that must be fulfilled for a phosphodiester bond to become a potential 3’ splice site. [1] It must be preceded by a G: We observed cleavage and ligation at a total of seven [possibly nine] locations, and all of them follow a G. [2] This G must not be sequestered in a double-stranded structure: Sites following G209 and G219, for instance, are not accessible unless the P9.1b pairing has been destabilized by several nucleotide changes. [3] The nucleotide preceding the G should base-pair with nucleotide 2 of the J7-9 segment, which is part of a newly described 3’SSBS common to many, if not all, group I introns: Both the U-A and C-G combinations at these sites were found to lead to efficient cleavage and ligation, the U-G, G-A, and A-A wobble and odd pairs allow only reduced activity, and the C-A pair allows almost none at all. Finally, for a well-matched 3’ splice site to gain access to the 3’SSBS, it must not be competed out by more proximally lying competitors: Once freed by destabilization of the P9.1b helix, the splice sites following G239 and G218, for instance, are not accessible unless the P9.1b pairing has been destabilized by several nucleotide changes.

In most members of subgroup IA [to which the sunY intron belongs], the sequence complementary to the
be a G and cannot be part of an RNA helix was already apparent both from sequence analyses and in vitro studies [Barford and Cech 1988; Price and Cech 1988]. Our identification of a binding sequence (within the intron) for the nucleotides immediately preceding the 3'-terminal G is new, although not unexpected: The efficiency of the circularization step in the Tetrahymena system had been shown to rely on tight binding of the 3' extremity of the intron [Tanner and Cech 1987]. Furthermore, mutating the two penultimate nucleotides of the intron reduces the efficiency of splicing, even in the context of an intact P10 pairing [Price and Cech 1988].

As seen in Figure 7, the most commonly encountered 3'SSBS consists of 2 nucleotides [either 1-2 or 2-3 of the J7-9 segment], making Watson–Crick base pairs pairing between nucleotides of 3'SSBS and penultimate intron shown. In-...
Michel et al.

tants in which the pairing is restored are partially back to normal [Michel et al. 1989].

Similarly, the sunY intron has two Us preceding its terminal G that can pair with the two As of its 3'SSBS. The fact that none of the splice junctions used by truncated transcripts of the sunY intron could make more than 1 bp with its 3'SSBS may have contributed to the reduced efficiency of proximal splicing and also to the ease with which we were able to change the specificity of 3' cleavage from UG to CG by changing only 1 nucleotide of the 3'SSBS (the first A to G). The first UUG that could serve as a 3' splice site occurs at G3269. 34 nucleotides 3' of P9.1, the last identifiable secondary structure element of truncated transcripts. None of the group I introns has a 3' splice site so distant from the structured core. The clear preference of the sunY core for proximal sites in truncated transcripts [note that the distance of the 3' splice site from the last structure element in full-length transcripts is 11 nucleotides—see Fig. 2a] may explain why a single base-pair interaction with the 3'SSBS is sufficient to determine the 3' splice site. Note also that the other two T4 introns [td and nddB], which are closely related to intron sunY (Shub et al. 1988), have only a 1-bp 3'SSBS interaction (Fig. 7).

A number of group I introns do not have a 2-nucleotide 3'SSBS but, rather, a longer or [more often] shorter one. In several introns, there is not even a single nucleotide in the J7-9 segment that could make a Watson–Crick base pair with the 2 nucleotides preceding the terminal G. Interestingly, these cases usually involve As on the proximal sites in truncated transcripts (note that the other two T4 introns [td and nddB], which are closely related to intron sunY (Shub et al. 1988), have only a 1-bp 3'SSBS interaction (Fig. 7).

We have recently observed another example where a shift in 3' splice site can be explained by interaction with the 3'SSBS. The 3' splice site of the nddB intron of phage T4 occurs between 2 G residues. The sequence at the 3' end of the intron is CGG, and splicing after the first G is consistent with base-pairing between the C and a G at position 2 of the J7-9 segment [Fig. 7]. We have recently described a hybrid intron, created by spontaneous deletion, that joins the core of td to the 3' end of nddB [Hall et al. 1989]. Splicing of this hybrid intron is displaced by 1 nucleotide relative to nddB, occurring after the second G, consistent with a C at position 2 of J7-9 of td [Fig. 7].

Other potential components of 3' splice selection, the P10 paradox, and alternative splicing

As first noted by Hensgens et al. [1983] and more recently discussed by Cummings et al. [1988], sites that mimic the intron–3' exon junction exist immediately downstream of the last core components of several mitochondrial introns [one case has recently been uncovered in Podospora anserina with 17 matches out of 20 residues, including all of the nucleotides destined to form the P10 stem as well as those to be recognized by the 3'SSBS (Cummings et al. 1990)]. These introns all include an ORF that cannot be translated from the 5' exon and that lies entirely 3' of their catalytic cores, both unusual features among group I introns of protein-coding genes. Moreover, alternative splicing to these sites would invariably put the ORF in phase with the 5' exon. Such coincidences, unlikely to have arisen by chance alone, have led to speculation that these sites support occasional alternative splicing events, leading to low-level expression of the intron-encoded protein [recall that all proteins encoded by mitochondrial introns are expressed at undetectably low levels in wild-type cells].

So far, we have been unable to demonstrate that this mechanism operates in the case of the sunY intron. Our experiments do show that truncated sunY transcripts undergo alternative splicing at high frequency in vitro. The paradox lies in the fact that the sites used are not the ones [G3273 and G3388] that would have been predicted assuming that the P10 pairing is the major determinant of the process of selecting a 3'-intron/exon junction. More generally, although most group I introns potentially include an extensive P10 pairing, in vitro experiments have long been unable to establish a clear proof of its existence [e.g., Been and Cech 1985]. As has now become apparent, one reason for this failure is that the 3' intron–exon junction seems to be specified in a redundant way. For disruption of the P10 pairing to show appreciable effects, one may need to destroy the 3'SSBS first. As has just been shown to be the case in the Tetrahymena intron [Michel et al. 1989], such a strategy should make it possible to assess the significance of potential P10 pairings in a variety of group I introns.

Experimental procedures

Plasmid construction and site-directed mutagenesis

The AccI–SspI fragment of bacteriophage T4 DNA, extending from positions 2747 to 3272 in the sequence of Tomaschewski and Rüger [1987], includes 26 nucleotides of the exon 5' to intron sunY, the entire core of the sunY intron (196 nucleotides, including structural elements P1–P9), as well as 304 nucleotides 3' of the Pvull site at the end of the P9 stem–loop structure. The AccI end of this fragment was made blunt with the Klenow fragment of Escherichia coli DNA polymerase I and ligated to the HindII site of plasmid pGEM-3 [Promega Biotech], 27 nucleotides from the SP6 transcription start [the SspI end of the fragment was ligated to the Smal site of the plasmid] to produce plasmid pGEM-3/T4AS. To allow transcription with the T7 RNA polymerase, this construct was transferred into plasmid pTZ19U [U.S. Biochemicals]. Site-directed mutagenesis was carried out either by the gapped duplex method of Inouye and Inouye [1987] for pGEM-3 derivatives or by the method of Kunkel et al. [1987] for pTZ19U derivatives.

In vitro transcription

Linearized plasmid DNA (2.5 μg) was incubated for 2 hr at 40°C with SP6 or T7 RNA polymerase in a buffer containing 10 mM NaCl, 40 mM Tris-HCl (pH 7.5), 6 mM MgCl2, 10 mM dithio-
threitol, 2 mM spermidine, 40 units of RNAsin [Promega Biotec], 1 mM of each of the four ribonucleotide triphosphates, and 20 μCi of [α-32P]GTP (400 μCi/m mole). No splicing could be detected during transcription under these conditions. Transcripts were purified by electrophoresis on a 4% acrylamide/8 M urea gel, followed by elution overnight in 200 mM Tris-HCl (pH 7.5), 225 mM Na2EDTA, 0.3 M NaCl, and 2% SDS.

Reverse transcription of splicing products
In vitro-transcribed RNA was ethanol-precipitated and resuspended in 50 mM Tris-HCl (pH 7.5), 50 mM NH4Cl, 30 mM MgCl2, 10 mM spermidine, 0.05% SDS, and 2 mM GTP and incubated at 37°C for 60 min. Splicing products were separated by gel electrophoresis and eluted as described above, except that carrier tRNA at a final concentration of 0.1 μg/ml was added. The eluted RNA was coprecipitated with 2 pmol of 32P-end-labeled oligodeoxynucleotide primer and resuspended in 60 mM NaCl, 50 mM Tris-HCl (pH 8.3) at 37°C and 10 mM dithiothreitol. The mixture was heated at 90°C and slow-cooled (30 min) to 30°C. Elongation was carried out with 6 units of Rous avian virus reverse transcriptase [Amersham] for 45 min at 40°C in the same buffer plus 40 mM MgCl2, 0.5 mM of each dNTP, and 0.1 mM of one of the four dNTPs. The reaction was stopped by precipitation with ethanol/sodium acetate.

Quantitation of 32P-labeled RNA species
Bands on autoradiographs were quantitated by optical scanning of varied exposures. The gel shown in Figure 6 was also directly counted on a Betascope two-dimensional array detector (Betagen, Waltham, MA), with results that were consistent with optical scanning.

Acknowledgments
We are indebted to Luc d'Auriol for providing us with oligonucleotides at a time of need; to Jack Szostak, in whose lab this work was completed while one of us (F.M.) was on sabbatical leave; and to John Jaeger and Doug Turner for checking on our calculations of helix stabilities. We thank Tom Cech and the referees for helpful comments on the manuscript. This work was supported by National Institutes of Health research grant GM-37746 to D.A.S.

References
Barfod, E.T. and T.R. Cech. 1988. Deletion of nonconserved helices near the 3' end of the RNA intron of Tetrahymena thermophila alters self-splicing but not core catalytic activity. Genes Dev. 2: 652–663.

Been, M.D. and T.R. Cech. 1985. Sites of circularization of the Tetrahymena rRNA IVS are determined by sequence and influenced by position and secondary structure. Nucleic Acids Res. 13: 8389–8408.

———. 1986. One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. Cell 47: 207–216.

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. 1988. Conserved sequences and structures of group I introns: Building an active site for RNA catalysis—A review. Gene 73: 259–271.

Chandry, P.S. and M. Belfort. 1987. 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. Genes Dev. 1: 1028–1037.

Cummings, D.J. and J.M. Domenico. 1988. Sequence analysis of mitochondrial DNA from Podospora anserina. Pervasiveness of a class I intron in three separate genes. J. Mol. Biol. 204: 815–839.

Cummings, D.J., J.M. Domenico, and F. Michel. 1988. DNA sequence and organization of the mitochondrial ND1 gene from Podospora anserina: Analysis of alternate splice sites. Curr. Genet. 14: 253–264.

Doudna, J.A. and J.W. Szostak. 1989. Miniregions, small derivatives of the sunY intron, are catalytically active. Mol. Cell. Biol. 9: 5480–5483.

Durocher, V., A. Gauthier, G. Bellemare, and C. Lemieux. 1989. An optional group I intron between the chloroplast small subunit rRNA genes of Clamydomonas moewiusii and C. eugametos. Curr. Genet. 15: 277–282.

Ewart, J.-L., M. Kunzt, N.A. Strauss, and J.-H. Weil. 1988. A class-I intron in a cyannele tRNA gene from Cyanophora paradoxa: Phylogenetic relationship between cyanneles and plant chloroplasts. Gene 71: 115–122.

Gott, J.M., A. Zeeh, D. Bell-Pedersen, K. Ehrenman, M. Belfort, and D.A. Shub. 1988. Genes within genes: Independent expression of phage T4 intron ORFs and the genes in which they reside. Genes Dev. 2: 1791–1799.

Hall, D.W., Y. Liu, and D.A. Shub. 1989. Exon shuffling by recombination between self-splicing introns of bacteriophage T4. Nature 340: 574–576.

Hensgens, L.A.M., L. Bonen, Muus de Haan, G. van der Horst, and L.A. Grivel. 1983. Two intron sequences in yeast mitochondrial COX1 gene: Homology among URF-containing introns and strain-dependent variation in flanking exons. Cell 32: 379–389.

Inouye, T., F.X. Sullivan, and T.R. Cech. 1986. New reactions of the ribosomal RNA precursor of Tetrahymena and the mechanism of self-splicing. J. Mol. Biol. 189: 143–165.

Inouye, S. and M. Inouye. 1987. Oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA. In Synthesis and applications of DNA and RNA (ed. S.A. Narang), pp. 181–206. Academic Press, Orlando.

Kunkel, T.A., J.D. Roberts, and R.A. Zakour. 1987. Rapid and efficient site-specific mutagenesis without phenotype selection. Methods in Enzymol. 154: 367–382.

Michel, F., A. Jacquier, and B. Dujon. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochemistry 21: 867–881.

Michel, F., M. Hanna, R. Green, D.P. Bartel, and J.W. Szostak. 1989. The guanosine binding site of the Tetrahymena ribozyme. Nature 342: 391–395.

Price, J.M. and T.R. Cech. 1988. Determinants of the 3' splice site for self-splicing of the Tetrahymena pre-rRNA. Genes Dev. 2: 1439–1447.

Price, J.V., J. Engberg, and T.R. Cech. 1987. 5' exon requirement for self-splicing of Tetrahymena thermophila pre-ribosomal RNA and identification of a cryptic 5' splice site in the 3' exon. J. Mol. Biol. 196: 49–60.

Shub, D.A., J.M. Gott, M.-Q. Xu, B.F. Lang, F. Michel, J. To-
maschewski, J. Petersen-Lane, and M. Belfort. 1988. Structural conservation among three homologous introns of phage T4 and the group I introns of eukaryotes. Proc. Natl. Acad. Sci 85: 1151–1155.

Sogin, M.L. and J.C. Edman. 1989. A self-splicing intron in the small subunit rRNA gene of Pneumocystis carinii. Nucleic Acids Res. 17: 5349–5359.

Tanner, N.K. and T.R. Cech. 1987. Guanosine binding required for cyclization of the self-splicing intervening sequence ribonucleic acid from Tetrahymena thermophila. Biochemistry 26: 3330–3340.

Tomaschewski, J. and W. Rüger. 1987. Nucleotide sequence and primary structures of gene products coded for by the T4 genome between map positions 48.266 and 39.166 kb. Nucleic Acids Res. 15: 3632–3633.

Tu, C., P. Gauss, C. Thermes, D.R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. d’Aubenton-Carafa, O.C. Uhlenbeck, I. Tinoco, Jr., E.N. Brody, and L. Gold. 1988. CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc. Natl. Acad. Sci. 85: 1364–1368.

Van der Horst, G. and H.W. Tabak. 1987. New RNA-mediated reactions by yeast mitochondrial group I introns. EMBO J. 6: 2139–2144.

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.

Xu, M.-Q. and D.A. Shub. 1989. The catalytic core of the sunY intron of bacteriophage T4. Gene 82: 77–82.

Note added in proof

After this manuscript was accepted for publication, we learned of the existence of an article that predicts the 3’SBS on the basis of an analysis similar to that presented in Figure 7. (J.M. Burke. Selection of the 3’-splice site in group I introns. FEBS Lett. 250: 129–133, 1989). That article assumes an invariant location of the 3’SBS (the first two nucleotides between P7 and P9) for all group I introns, whereas we have made distinctions between the various subgroups. The Burke analysis does not explain the preference for penultimate U over C by the sunY intron core, nor would it have predicted the change in 3’ splice site selection caused by the A-177 to G substitution.
Mechanism of 3' splice site selection by the catalytic core of the sunY intron of bacteriophage T4: the role of a novel base-pairing interaction in group I introns.

F Michel, P Netter, M Q Xu, et al.

*Genes Dev.* 1990, 4:
Access the most recent version at doi:10.1101/gad.4.5.777

---

**References**
This article cites 30 articles, 7 of which can be accessed free at:
http://genesdev.cshlp.org/content/4/5/777.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.