Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron self-splicing

Sulayman D.Dib-Hajj⁺, Scott C.Boulanger¹, Sharda K.Hebbar, Craig L.Peebles², James S.Franzen² and Philip S.Perlman^{1,*}

¹MCDB Program and Department of Molecular Genetics, The Ohio State University, 484 W.

12th Avenue, Columbus, OH 43210, ¹Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-9038 and ²Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received January 25, 1993; Revised and Accepted March 18, 1993

ABSTRACT

The role of domain 5 (d5) from the self-splicing group II intron 5 γ of the *COXI* gene of yeast mitochondrial DNA in branching and 3' splice site utilization has been studied using a substrate transcript lacking d5 (Δ d5 RNA). This RNA is completely unreactive *in vitro*, but releases 5' exon by hydrolysis under various reaction conditions when d5 RNA is added in *trans*. Under an extreme reaction condition, some accurate branching and splicing occur. Much more efficient use of a 3' splice site is obtained when Δ d5 RNA is complemented by a transcript containing the wild-type domains 5 and 6 plus the 3' exon. While most Δ d5 RNA molecules in that protocol still react by hydrolysis at the 5' splice site, the branching that occurs uses only the d6 tethered to d5 that is provided in *trans*. The use of this d6 and the 3' splice site also linked to d5, along with the observed indifference to the other d6 and 3' splice site resident in the Δ d5 RNA, indicates that d5 plays a key role in positioning d6 for the first reaction step as well as in 3' splice site use. Two models for the manner by which d5 interacts with d6 are discussed.

INTRODUCTION

Group II introns are distinguished from group I introns, the other major class of self-splicing introns, by their distinctive conserved secondary structure consisting of six substructures or domains (1, 2). A number of short conserved primary sequences are present in the various domains, including domain 5 (d5) itself, which is the most highly conserved primary sequence element in group II introns. The typical d5 structure contains 34 nts and is predicted to fold as a hairpin. The hairpin may be described as an extended 14 basepair helix, capped by a four base loop involving d5 positions 15–18, and punctuated by a 2 base bulge at positions 25 and 26.

In vitro trans-splicing of the self-splicing group II intron 5 γ of the yeast *COXI* gene (aI5 γ) was previously demonstrated using individually unreactive fragments of the original self-splicing transcript (3). The bimolecular reaction was shown to require the d5 sequence within the 3' half molecule. An RNA containing primarily the d5 sequence was sufficient to restore 5' exon (5'E) release from a 5' half molecule containing the 5'E plus intron domains 1, 2, and 3 (5'E-d123). That finding was interpreted to mean that d5 makes a tertiary contact with some site(s) in 5'E-d123 and that the interaction is needed to permit any reaction at the 5' splice site to occur. No branching or splicing was restored when an RNA containing domain 6 (d6) and the 3' exon (3'E) was provided as a third reactant. Based on that result, it was suggested that d5 may also play some role in the second splicing step, either selection and/or cleavage at the 3' splice site. We have recently analyzed a derivative of aI5 γ deleted for domains 2, 3, 4, and 6 and found that the mini-intron retains the 5'E release reaction (4). That result showed that the binding site for d5 involves sequences primarily within domain 1.

Several studies of 3' splice site selection in group II introns have appeared (5–9). They have identified two weak contacts that play roles in defining the 3' splice site. The first of these contacts is a lone basepair termed the γ/γ' interaction between the 3' terminal nucleotide of the intron and a single base within junction 3, the sequence between d2 and d3 (7). A second single basepair interaction, termed the internal guide, has been defined between the first base of the 3' exon and the nucleotide adjacent to the 5' end of the exon binding site (EBS1) element in domain 1 (9). Several groups have reported mutations that alter 3' splice site selection, including the deletion of d6 (4–8). However, those studies did not explore the involvement of d5 in reactions at the 3' splice site. Here we examine *trans*-splicing reactions of d5 from aI5 γ that provide strong support for the hypothesis that d5 plays a significant role in positioning d6 and the 3' splice site for branching and splicing.

* To whom correspondence should be addressed

⁺ Present address: Department of Entomology, Texas A&M University, College Station, TX 77843, USA

MATERIALS AND METHODS

Plasmids

Plasmid pJD20 (3) was used as template for transcription of control precursor RNA; it contains a fragment of mtDNA (EMBL accession number JO 1481, nucleotides 8477-9923) containing the wild-type *aI5 γ* plus flanking exon sequences from yeast strain D273-10B. Plasmid pJD20 Δ d5 is a derivative of pJD20 (4) that was used as template for making a precursor RNA transcript deleted for 32 nt of d5 (Δ d5 RNA); residues 9561-9572 of the cited wild-type sequence are deleted. Plasmid pJDI5'-75 was the template for d5 RNA and an RNA containing d5, d6 and 3'E sequences (d56-3'E) (3).

Transcription and *in vitro* splicing

RNAs were made by transcribing linearized plasmids using T7 RNA polymerase, in most cases in the presence of 20 μ Ci of α -³²P-UTP (Amersham) as described previously (3). To obtain precursor RNAs containing the shorter version of 3'E (3'E^S), pJD20 or pJD20 Δ d5 were linearized by cleavage with HindIII; to prepare precursor RNAs containing the longer version of 3'E (3'E^L), the plasmids were cleaved with PvuII; this treatment adds 226 nt of vector sequence to the 3' end of the 3' exon. To obtain d5 RNA, pJDI5'-75 was cleaved with HpaII; and to obtain d56-3'E^S or d56-3'E^L RNAs, pJDI5'-75 was cleaved with HindIII or PvuII, respectively. All of the *in vitro* splicing reactions contained 100 mM MgCl₂ and 40 mM HEPES (pH 7.5); some also contained 0.1% SDS, though that compound does not influence these self-splicing reactions. Additions of KCl, NH₄Cl, or (NH₄)₂SO₄ were included at final concentrations of 0.5 M or 1.5 M, as noted in the legends and text. In complementation experiments, both RNAs were mixed and the reaction was started by addition of the reaction buffer. All reactions were carried out at 45°C for the times noted. Incubations were terminated by addition of 1 volume of gel loading buffer and samples were loaded directly onto 4% polyacrylamide gels (19/1 acrylamide to bisacrylamide) containing 8 M urea. Gels were dried and analyzed by autoradiography or with the Molecular Dynamics PhosphorImager. Abbreviations of various RNA species detected on gels are defined in the figure legends or in the text.

Other biochemical methods

Debranching using HeLa cell S100 extracts was carried out essentially as described by Ruskin and Green (10). Isolated labeled RNA (5000 cpm) was analyzed in 25 μ l reactions having a final concentration of 20 mM HEPES (pH 8), 10 mM EDTA with various amounts of extract as noted. Samples were incubated for 1 hr at 30°C. Reactions were terminated according to Arenas and Hurwitz (11) and analyzed on gels as above. The control RNAs for this experiment were purified using an RNAid kit (Bio 101, Inc.). Because the lariat RNAs purified from the *trans*-reactions were already relatively salt-free, it was not necessary to purify them using RNAid. RNase T1 digestion experiments were conducted as described previously (12). cDNA formation using spliced exon RNAs as template and PCR amplification of the cDNA product were carried out as described in ref. 4. Dideoxy chain termination DNA sequencing was done using a Sequenase kit (US Biochemical) or a fmol[™] kit (Promega), according to instructions.

RESULTS

Complementation of Δ d5 precursor RNA by addition of d5 RNA

We have previously described *trans* reactions of d5 from *aI5 γ* using a substrate RNA containing the 5'E plus intron domains 1, 2, and 3 (5'E-d123) (3). That study demonstrated that d5 is essential for the first splicing step, the release of 5'E RNA. However, since the 3' splice site is absent from that substrate RNA, no insight could be obtained about the possible role of d5 in the second splicing step. In the present study, we used a derivative of the wild-type *aI5 γ* transcript lacking d5 (Δ d5 RNA; later denoted as Δ d5-3'E^S) that permits us to examine d5-dependent *trans* reactions of *aI5 γ* where both splice junctions are present within the same substrate RNA. The Δ d5 RNA is completely inactive under all reaction conditions that we have tested (see ref. 4), including those with concentrations of added salt as high as 1.5 M.

Adding d5 RNA in *trans* complements Δ d5 RNA in a number of high salt conditions (see Fig. 1, lanes 4–9). Some reaction is observed in 1.5 M (NH₄)₂SO₄ (lane 5), but there is none in 0.5 M (NH₄)₂SO₄ (lane 4); the main products are linear IVS(Δ d5)-3'E and 5'E RNAs, resulting from hydrolysis at the

Figure 1. *Trans* complementation of Δ d5 RNA by domain 5 RNA. Δ d5 RNA and d5 RNAs were prepared by transcription and analyzed for *in vitro* reactions in various buffers. All samples contain a final concentration of 100 mM MgCl₂ and 40 mM HEPES (pH 7.5); (NH₄)₂SO₄, NH₄Cl or KCl were added to the final concentrations noted in the figure. Complementation reactions in lanes 4–9 contained about 700 nM d5 RNA and about 30 nM Δ d5 RNA and were incubated for 45 min at 45°C. Lanes 1 and 3 contain unreacted d5 and Δ d5 RNA, respectively; reaction of each of those RNAs yields no products (not shown). Lanes 2 and 10 contain control *aI5 γ* precursor RNA incubated for 15 min in buffer containing 0.5 M NH₄Cl. The IVS(Δ d5)-LAR product is observed only in lane 7. The asterisk (*) denotes a d5-dependent product of cleavage of intron RNAs lacking 5' E (3, 12, 13). Abbreviations used to refer to precursor and product RNAs are the following: IVS-LAR, intervening sequence (intron) lariat; TXT, Δ d5 or wild-type precursor RNA, as indicated; IVS(Δ d5)-3'E, reaction intermediate containing Δ d5 intron and 3'E; IVS-LIN, wild-type linear excised intron; IVS(Δ d5), linear excised Δ d5 intron; 5'E-3'E, spliced exon; 3'E, 3' exon; 5'E, 5' exon.

5' splice junction. Reactions containing NH_4Cl are significantly more efficient (lanes 6 and 7) and in 1.5 M NH_4Cl some reaction even occurs at the 3' splice site. For example, both 3'E and excised linear intron RNAs are evident in lane 7. The ratio of 5'E to 3'E products shows that only about 20% of molecules that react at the 5' splice site also splice. Under that reaction condition even the wild-type precursor RNA carries out the spliced exon reopening (SER) reaction so that free 5' and 3' exon RNAs accumulate instead of spliced exon RNA. Even though spliced exon RNA is not a prominent product of this reaction, we tested the accuracy of splicing by amplifying the spliced exon RNA using reverse transcriptase and PCR. The amplified cDNA was cloned and each of two splice junctions sequenced resulted from correct splicing.

Some branching occurs in 1.5 M NH_4Cl (see lane 7) since some excised Δd5 intron lariat RNA is clearly present. RNase T1 analysis of that lariat RNA shows that it probably contains the correct branch structure involving residue A880 of the d6 substructure and the first intron base (not shown, but see ref. 3, 15 for similar experiments). Since about 90% of the excised Δd5 intron RNA and all of the IVS(Δd5)-3'E RNA is linear, it is clear that the *trans*-acting d5 RNA restores branching much less efficiently than splicing. Interestingly, the branched intermediates appear to be more likely to undergo the second step of splicing, because there is no detectable lariat intermediate (intron plus 3' exon) while the linear intermediate accumulates abundantly. It is possible that some 5'E results from d5-dependent cleavage of incomplete transcripts or broken molecules, i.e., from reactions analogous to the complementation between 5'E-d123 and d5; however, since there is a prominent yield of IVS-3'E in these experiments it is clear that the reaction involving incomplete Δd5 transcripts is quantitatively minor.

Figure 2. Diagram of transcripts used to analyze complementation with splicing. The schematic diagram illustrates the precursor RNAs used in the experiments of Figures 3, 4, and 5. It defines a modified nomenclature of the precursor and product RNAs in which two different length 3'E sequences are distinguished. These experiments were designed to test whether Δd5 -3'E^L RNA can interact with d56-3'E^S RNA to release 5'E and whether Δd5 -3'E^L RNA can also carry out branch formation and/or splicing with either or both of the 3' splice junctions present. While the 3' splice site sequences are identical on both RNAs, the 3'E elements differ in length by 216 nts as indicated. The extra sequences are derived from the vector and appear at the 3' end of 3'E^L only.

In the control reaction with full-length self-splicing transcript incubated in 0.5 M NH_4Cl (lanes 2 and 10), branching occurs in the first reaction step at least 80% of the time. There is only a little branching in the d5-promoted reactions of Δd5 RNA in lane 7 and there is no (or very much less) branching under all of the other conditions tested here. Thus, the considerable defect in branching during the first reaction step probably results from a defect in positioning d6, which contains the branchpoint nucleotide.

Complementation by added d5 RNA is also observed in reactions containing KCl (Fig. 1, lanes 8 and 9). The main products are 5'E and linear IVS(Δd5)-3'E RNAs, although there is a slow reaction at the 3' splice junction. We have previously examined a number of mutant versions of $\text{d5}\gamma$ that are blocked for the reaction at the 3' splice site in 0.5 M KCl but regain activity there when the KCl concentration is raised to 1.5 M (4). Here, higher KCl yields a somewhat greater rate and extent of reaction at the 5' splice site and activates the 3' splice site somewhat; however, the activation of the 3' splice site by 1.5 M salt was much more obvious with NH_4Cl . Using our previous *trans*-splicing system, where each exon is provided by a separate half molecule, comparable extents of reaction are obtained at both splice junctions under all of these reaction conditions (3, 13). Therefore, we conclude that the reduced ability of Δd5 RNA to carry out an efficient reaction at the 3' splice site under all salt conditions when d5 is provided in *trans* probably reflects a role for d5 in the second reaction step.

Franzen et al. (14) have recently published measurements of the K_m for d5 in several complementation reactions, one of which is the same as used here. They report a K_m of nearly 300 nM and note that it varies somewhat with changes in salt concentration. They used a d5 RNA that is 36 nt long while we used a longer version here (the same one used by Jarrell et al. (3) in the original report of d5 complementation). We have also examined the dependence of the complementation reaction on the concentration of the longer d5 RNA; that experiment was not conducted with the intention of measuring a K_m , but it demonstrated that the level of d5 used here was at or above the amount needed for a substantial extent of reaction in a short incubation.

Domain 5 plays a role in positioning the 3' splice junction

As shown above, the interaction between Δd5 and d5 RNAs leads mainly to release of the 5'E RNA. We have already reported that 5'E-d123 RNA *trans*-splices when incubated with d56-3'E RNA in 0.5 M $(\text{NH}_4)_2\text{SO}_4$ and 0.5 M KCl (3); *trans*-splicing also occurs in 0.5 M NH_4Cl (13). Consequently, we tested whether Δd5 RNA can execute the second splicing step more efficiently when the 3' splice junction is provided together with d5 and d6 (as the d56-3'E RNA). If d6 and/or 3'E bind to other regions of the intron independently of d5, then those sites should be occupied in *cis* in this experiment. In that case, the d6 and 3'E of the Δd5 RNA might block access to the reaction center for d56-3'E RNA. On the other hand, if the binding of d5 of the d56-3'E RNA positions the 3' splice site on the intron appropriately, then the d6 and 3'E provided in *trans* should be used efficiently and preferentially over the d6 and 3'E provided by the Δd5 RNA. The latter prediction also applies if the putative d6 or 3'E binding sites do not form on the Δd5 RNA until d5 RNA is bound. Because there are two potential sources of the

Figure 3. *Trans* complementation of $\Delta d5$ RNA with efficient splicing. Transcripts $\Delta d5$ -3'E^L and $d56$ -3'E^S were labeled to the same specific radioactivity, mixed to give a molar ratio of about 1/3 (170 nM $d56$ -3'E^S and 50 nM $\Delta d5$ -3'E^L), and incubated in standard reaction buffer containing 1.5 M NH_4Cl at 45°C for the times indicated above lanes 3–5. Lane 2 is $\Delta d5$ -3'E^L RNA incubated alone in the same buffer for 45 min. Lanes 1 and 6 are control (CON) self-splicing reactions of wild-type transcripts containing 3'E^S and 3'E^L, respectively. Note the distinct mobilities of the spliced exons with different 3'E elements and of the free 3'E products that differ in length by 216 nts. Lanes 8–10 represent the same experiment as lanes 3–5 except that only the $d56$ -3'E^S has been labeled. Lane 7 is $d56$ -3'E^S RNA incubated alone in the same buffer for 45 min. Lanes 11–14 are a fifty-fold longer exposure of lanes 7–10, shown to reveal slowly migrating, low-abundance products of the complementation reaction. Here and subsequently, all RNAs containing a 3'E element are designated according to whether 3'E^L or 3'E^S is present; see Figure 2 for a definition of how the two 3'E sequences differ. Abbreviations not defined earlier are: Y-IVS, Y-branched intron RNA containing $\Delta d5$ IVS and $d56$ RNAs joined by a 2'-5' phosphodiester branch; Y-IVS-3'E^L, Y-branched reaction intermediate containing $d56$ joined to IVS($\Delta d5$)-3'E^L intermediate. Figure 7 illustrates all of these RNA species in diagrammatic form.

3'E in this experiment, one linked to $d5$ and another linked to the rest of the intron plus the 5'E in the $\Delta d5$ RNA, we used 3'E elements of different lengths. The $\Delta d5$ RNA used in this experiment extends to the plasmid Pvu II site to create a 3'E of 532 nts (3'E^L); the $d56$ -3'E RNA ends at the polylinker HindIII site to produce a 3'E of 316 nts (3'E^S) as shown schematically in Fig. 2.

Preliminary experiments showed that the $\Delta d5$ -3'E^L RNA actively releases both 5'E and IVS($\Delta d5$)-3'E^L products when incubated in high salt buffers with $d56$ -3'E^S RNA. The 3'E^S, which is linked to $d5$, is also released efficiently under all of the conditions described above as active for $\Delta d5$ RNA complementation; here splicing followed by SER accounts for the free exon RNAs. A typical experiment using 1.5 M NH_4Cl is shown in Fig. 3; essentially the same result was obtained in buffers containing 0.5 and 1.5 M KCl, 0.5 M NH_4Cl and 0.5 and 1.5 M $(\text{NH}_4)_2\text{SO}_4$. Control lanes define the location of spliced and free exon RNAs derived from self-splicing wild-type transcripts terminated at the different plasmid sites; 3'E^S and 3'E^L are prominent products in the control lanes 1 and 6, respectively. Lanes 4 and 5 show that reacting $d56$ -3'E^S and $\Delta d5$ -3'E^L RNAs yields both 5'E and 3'E^S RNAs as products. There is also an inefficient reaction at the 3' splice site of $\Delta d5$ -3'E^L RNA, since there is a little linear $\Delta d5$ excised intron RNA. In a longer exposure of this gel, some 3'E^L is also seen. However, there is no indication of any use of the $d6$ branch site

Figure 4. Kinetics of accumulation of slowly migrating, low-abundance products of complementation by $d56$ -3'E^S. An experiment similar to that of Fig. 3, lanes 7–10 was performed with reaction times of 0, 15, 30, 45, 90, and 180 min. The resulting gel was dried and analyzed as before with the PhosphorImager. The figure displays the relative radioactivity of several slowly migrating products. The total radioactivity varied from lane to lane by less than 10%, so the data shown were not corrected for variations in loading. Each data set is labeled to identify the RNA species described according to the abbreviations used in Fig. 3. In this experiment, Y-IVS-3'E^LS, a more slowly migrating branched product containing both 3'E^L and 3'E^S, was detected. The amount of radioactivity in each species may be converted to a molar representation by adjusting for their relative UMP content. For example, the species Y-IVS-3'E^LS, Y-IVS-3'E^L, Y-IVS, and 5'E-3'E^S are present in a molar ratio of 1/63/100/9 at the 45 min time of reaction.

in the $\Delta d5$ -3'E^L RNA, since no intron lariat or IVS($\Delta d5$)-3'E^L lariat RNAs appear (lanes 4 and 5, and longer exposures, not shown).

The ratio of 5'E to 3'E^S produced by the *trans* reaction (Fig. 3, lanes 4 and 5) is very similar to the ratio observed in reactions of the wild-type transcript (lane 1). This result shows that the 3' splice site linked to $d5$ is nearly as reactive as the 5' splice site present on the $\Delta d5$ RNA. Thus, the 3' splice site and $d6$ elements present in the $\Delta d5$ RNA are essentially excluded from the reaction. These data provide strong evidence that $d5$ plays a direct role in positioning and selecting the 3' splice junction. It is unlikely that the $d6$ and 3'E^L of the $\Delta d5$ RNA occupy $d5$ -independent sites on the intron, since their exogenously added counterparts are preferentially reactive.

A minor *trans*-splicing pathway involves branching

Additional information was obtained from a second version of this complementation experiment in which radioactively labeled $d56$ -3'E^S RNA was reacted with unlabeled $\Delta d5$ -3'E^L RNA (Fig. 3, lanes 7–10). In a short exposure (lanes 7–10), it is evident that a major product is 3'E^S RNA. This finding confirms that the product RNA in lanes 4 and 5 that comigrates with the 3'E^S marker (lane 1) is derived from the $d56$ -3'E^S RNA. The other principal labeled product obtained in this experiment is $d56$ RNA. This gel was run for a long time to enhance the resolution of the most slowly-migrating products, so that the $d56$ product was run off (see Fig. 6, lane 8 for an example where $d56$ RNA is shown). Scanning of a replicate of this experiment in which $d56$ was retained on the gel reveals that over 90% of $d56$ -3'E^S molecules that react yield those two products in a 1:1 ratio. Since most $d56$ RNA is not involved in

Figure 5. Identification of products containing 3'E sequences from d56-3'E RNA. Lanes 1–3 contain unlabeled $\Delta d5-3'E^L$ RNA and labeled d56-3'E^S RNA incubated in 1.5 M NH_4Cl for 0, 30 and 60 min., respectively. Lane 4 contains a control (C) wild-type transcript containing 3'E^S, incubated in 1.5 M NH_4Cl buffer for 5 min. Lanes 5–7 contain unlabeled $\Delta d5-3'E^L$ RNA and labeled d56-3'E^L RNA incubated as in lanes 1–3. Lane 8 contains control transcript containing 3'E^L, incubated in 1.5 M NH_4Cl for 5 min. As in Figure 3, this gel was exposed long enough to detect the slowly migrating branched and spliced exon products of the experimental samples.

branched structures, it follows that the first splicing step in the main reaction pathway occurs by hydrolysis instead of transesterification (see Fig. 7).

Longer exposures (Fig. 3, lanes 11–14) reveal the products of the less efficient branching and splicing reactions that transfer radioactivity from the relatively fast-migrating d56-3'E^S RNA to slower migrating product RNAs containing some or all of the unlabeled $\Delta d5-3'E^L$ RNA. Lanes 13 and 14 show three labeled products migrating slower than the input labeled d56-3'E^S RNA. One product (5'E-3'E^S) migrates with spliced exon RNA containing 3'E^S (compare lanes 13–14 with lane 1), showing that some splicing of 5'E to 3'E^S occurs. That material was extracted from a gel slice, converted to a cDNA and PCR amplified; sequencing of the amplified material showed that most, if not all, spliced molecules have the wild-type splice junction. Our previous studies of the spliced exon reopening (SER) reaction that occurs under these reaction conditions show that the 3'E^S RNA results from splicing followed by hydrolysis of the bond connecting 5'E and 3'E sequences (rather than hydrolysis of the 3' splice junction in the IVS-3'E reaction intermediate (4, 12).

A second product (Y-IVS-3'E^L) migrates near the linear IVS($\Delta d5$)-3'E^L product; Y-IVS-3'E^L is probably a branched, Y-shaped RNA. Based on its mobility, Y-IVS-3'E^L probably contains unlabeled IVS($\Delta d5$)-3'E^L as well as radioactive d56 from the complementing RNA. A third product (Y-IVS) migrates a little slower than the full-length linear IVS. Y-IVS must be branched to d56, because it has become labeled, but it appears to lack 3'E^L, based on its mobility; Y-IVS probably results from release of 3'E^L from Y-IVS-3'E^L by a second round of splicing (see Fig. 6). The second round of splicing presumably recycles 5'E derived from SER that follows the first splicing event. The second round of splicing is also followed by SER to account for the appearance of small amounts of 3'E^L that can be seen more clearly in an intermediate exposure of lanes 4 and 5 (not shown). Unlabeled material migrating at the position of 5'E-3'E^L was

Figure 6. Debranching of purified Y-IVS product RNA. The experiment of Fig. 3, lanes 2–5 was repeated except that the d56-3'E^S transcript was prepared with 16-fold higher specific radioactivity than the $\Delta d5-3'E^L$ RNA. The Y-IVS species was isolated after gel electrophoresis, and an untreated sample of Y-IVS is run in lane 3. Aliquots of Y-IVS were incubated with 0, 1, 5, or 10 μl of HeLa cell S100 extract (lanes 4–7). Controls include unreacted d56-3'E^S (lane 9) and d56-3'E^S reacted with unlabeled $\Delta d5-3'E^L$ RNA in 1.5 M NH_4Cl for 30 min (lane 8). The products of that reaction are indicated beside lane 9; note that the d56 RNA product is detected in lane 8. The mixture of products formed by a 1.5 M NH_4Cl reaction of the wild-type $\alpha 5\gamma$ self-splicing transcript is shown (lane 2). The product mixture is shown after debranching with 5 μl of HeLa cell S100 extract (lane 1) leading to a significant depletion of IVS-LAR; in that lane there is also an associated increase of linear intron RNA (that comigrates with Y-IVS) that is quite evident in a lighter exposure (not shown). A 5-fold longer exposure of the d56 region of the gel is shown below the main panel to reveal the d56 products more clearly. The Y-IVS fraction yields only a little d56 RNA because much of the purified material is labeled IVS-LIN.

analyzed as before, showing that the second round of splicing is also accurate.

We next measured the kinetics of accumulation of these minor products of *trans*-splicing that were discussed above. As shown in Fig. 4, the product identified as Y-IVS-3'E^L appears rapidly, reaches a maximum at about 45 min, and then declines after longer incubation. The product Y-IVS first appears after a 15 min delay and then accumulates throughout the rest of the time course. The product 5'E-3'E^S appears quickly and then decreases. In this experiment (and several others, but not the one shown in Fig. 3) a fourth minor band was evident. It has a mobility intermediate between those of lariat and precursor RNAs in the control lane and is deduced to be Y-IVS-3'E^{LS}, the branched intermediate containing both 3'E^S and 3'E^L formed in the first reaction step. As shown in Fig. 4, it is most evident at the early time points and then declines.

Analysis of products of the branching pathway of *trans*-splicing

To confirm our inferences regarding the 3'E content of these products, we carried out the same experiment with labeled d56-3'E^L and unlabeled $\Delta d5-3'E^L$ (Fig. 5). The rationale is that any product containing the 3'E from the labeled RNA would be

Figure 7. Proposed pathway of *trans*-splicing with $\Delta d5$ -3'E^L and $d56$ -3'E^S RNAs. The left side of the diagram labeled 'Major Pathway' outlines the reaction pathway taken by the majority of $\Delta d5$ -3'E^L and $d56$ -3'E^S RNAs that react in 1.5 M NH_4Cl . The principal products are free 5'E, 3'E^S, $d56$, and linear IVS-3'E^L. The right side of the diagram, labeled 'Minor Pathway', outlines the pathway that yields branched intermediates and products. In the first step, the branch site on $d56$ -3'E^S forms a branch with the first intron base of $\Delta d5$ RNA forming Y-IVS-3'E^{L-S} and 5'E RNA. The second step is splicing at 3'E^S to form spliced exons and Y-IVS-3'E^L. Spliced exon reopening can be catalyzed by the branched intron product (3). Y-IVS-3'E^L is converted efficiently to Y-IVS by a second round of splicing using 5'E RNA to produce some 5'E-3'E^L RNA. That minor spliced exon RNA is then converted to separate exons by a second round of SER. With the exception of the 5'E-3'E^L RNA, all of these products and intermediates have been illustrated in at least one figure.

shifted to a slower mobility, while molecules lacking that 3'E would be unaffected. As expected, the product identified as the spliced exon RNA (lanes 2 and 3) shifted to comigrate with spliced exons from the PvuII runoff transcript (lanes 6 and 7 vs lane 8). The products Y-IVS and Y-IVS-3'E^L were evident in this experiment, and exhibited the same mobility as before, thus confirming that they lack the 3'E provided in *trans* (compare lanes 2 and 3 with 6 and 7). And finally, scanning the top of this gel detected a shifted version of the Y-shaped intermediate containing both 3' exons. Together, these results clearly define those slowly-migrating products containing the 3'E provided by the $d56$ -3'E RNA.

In order to verify our identification of several branched products, we treated those RNAs with a HeLa cell S100 fraction that contains a debranching activity known to hydrolyze the 2',5'-linkage of group II intron branches (12, 15), including Y-branches resulting from *trans*-splicing (3). In trial experiments, we found that it was difficult to achieve complete debranching if the branched molecules were rare species in complex mixtures of reaction products. However, debranching could be readily demonstrated using the unfractionated products of self-splicing reactions in which IVS-LAR is a major product (Fig. 6, lanes

1 and 2). It was therefore necessary to enrich each branched RNA prior to debranching; but, since little radioactivity was present in the branched RNAs from the protocol used in Fig. 3, it was difficult to identify and purify the correct bands. To circumvent this problem, we carried out *trans*-splicing with both $d56$ -3'E^S and $\Delta d5$ -3'E^L labeled, but with $d56$ -3'E^S at 16-fold higher specific radioactivity than $\Delta d5$ -3'E^L. This modified protocol increased the radioactivity in the region of the gel containing the branched products and simplified their identification and isolation.

The purified RNA fraction containing Y-IVS (and linear intron RNA) was treated with HeLa S100 extract and fractionated on a gel (Fig. 6, lanes 5–7). It is clear that debranching releases material from the Y-IVS that comigrates with $d56$ RNA (compare lanes 5–7 with lane 8). There is no radioactivity migrating at that position in the untreated or mock-incubated samples (lanes 3 and 4), while efficient debranching of the control lariat does not release any small RNA product (lane 1). This experiment confirms that the major slowly-migrating product from these *trans*-splicing reactions is branched and contains radioactive $d56$ RNA. Since debranching did not release any $d56$ -3'E^S RNA or any other discrete product except $d56$ itself, we conclude that Y-IVS lacks the 3'E^S element. A similar experiment with Y-IVS-3'E^L confirms our identification of that product (not shown); it, too, released only $d56$ RNA as a labeled band after debranching. A reaction pathway consistent with these demonstrated structures and the kinetics of appearance and disappearance of the several products is provided in Fig. 7 and discussed below.

DISCUSSION

In the present study we have analyzed a derivative of the self-splicing group II intron $\alpha I5\gamma$ from yeast mtDNA in which all of the components of the precursor RNA, except for $d5$, are on a single transcript ($\Delta d5$ RNA). That molecule is completely unreactive *in vitro* but regains a significant 5' splice junction hydrolysis reaction upon incubation with $d5$ RNA under a variety of reaction conditions. In such complementation experiments there is only a little branching, consistent with our earlier suggestion that $d5$ may deliver $d6$ to the splicing reaction center (3). We found one reaction condition in which correct branching occurred even though $d5$ and $d6$ are on separate transcripts; this implies that the interaction between those two helices can be achieved in *trans*, albeit inefficiently. Under that condition complementation of the second splicing step was obtained but it is clear that only a fraction of molecules that splice also branched in the first reaction step. Current knowledge of 3' splice site selection points to roles for several specific nucleotides that form weak tertiary contacts but none of those interactions implies a role for $d5$ (7, 9).

The complementation experiment in which $\Delta d5$ RNA is reacted with $d56$ -3'E RNA differs from our earlier published study of *trans*-splicing (3) in that there is a choice of $d6$ and 3' splice sites. We found that the $d6$ tethered to $d5$ was used in branching to the apparent exclusion of the second copy of $d6$. *Trans*-branching is observed under a broader range of reaction conditions in this configuration than in the $\Delta d5$ plus $d5$ experiments. That survey of reaction conditions includes several in which $d5$ complementation reactions yield no branched products. However, the extent of branching was not as high as in the wild-type reaction (or as in *trans* splicing experiments

where each exon and intron domain is present only once), probably because of some interference in forming the optimal structure due to the presence of two copies of d6 and 3'E sequences.

These new data also support our earlier suggestion that d5 may be important for the second step of splicing (3, 13). It was interesting to find that removal of d5 from the self-splicing transcript does not necessarily incapacitate the 3' splice site. In complementation experiments in buffer containing 1.5 M NH₄Cl (Fig. 1, lane 7), about 20% of molecules that have reacted at the 5' splice site progress through the second splicing reaction. Linear IVS(Δ d5)-3'E is a major product of that reaction but little or no lariat IVS(Δ d5)-3'E accumulates, even though excised intron lariat RNA is quite evident. This situation implies that molecules that branch almost always complete the splicing reaction.

A vigorous reaction occurs at the 3' splice site of d56-3'E RNA when it is added in *trans* to the Δ d5 RNA. The near parity of the reactions at the 5' splice site of Δ d5 RNA and at the 3' splice site of d56-3'E RNA, even though most molecules react without branching, suggests that the 3' splice site is delivered to the reaction center after the complementing RNA was bound, presumably via d5. We showed previously that a precursor RNA deleted for d6 splices rather efficiently but with reduced fidelity at the 3' splice site (4). The finding that some correct splicing occurs even when d6 is absent shows that the interaction between d5 and d6 is not essential for the delivery of the 3' splice site to the reaction center. In an experiment in which Δ d5 RNA was complemented by a derivative of d56-3'E RNA lacking d6, preferential (but inaccurate) use of the 3' splice site tethered to d5 was still obtained (not shown).

As summarized in Figure 7, most molecules splice using the 3' splice site of the d56-3'E molecule, but without branching. A relatively minor pathway involves branching in the first step using the d6 that is linked to d5 to form the very minor intermediate Y-IVS-3'E^{LS} (containing both 3' exons). That molecule appears to be very reactive, splicing using only the 3' exon from the d56-3'E^S molecule (yielding Y-IVS-3'E^L and spliced exon RNA). Significantly, the alternative branched intermediate, Y IVS-3'E^S, was not detected, indicating that the first round of splicing is virtually limited to the 3' splice site tethered to .56. The absence of Y-IVS-3'E^S, which would migrate between Y-IVS-3'E^L and Y-IVS in lanes 13 and 14 of Fig. 3, is especially noteworthy since it would be detected with much greater sensitivity than either of the two branched molecules present there (because it contains many radioactive UMP residues from the 316 nt long 3'E^S).

We were surprised to find that Y-IVS (rather than Y-IVS-3'E^L) is the most abundant branched RNA present at late time points (see Fig. 4). Our data indicate that Y-IVS results from a *second* round of splicing using 5'E RNA produced after the first round of splicing by the spliced exon reopening (SER) reaction. We detected the predicted spliced exon RNA from that reaction (5'E-3'E^L) and found that splicing was accurate. Importantly, this reaction is secondary to the reaction using the 3' splice site of d56-3'E^S. It is curious that Y-IVS-3'E^L can splice a second time, since the d6 near the second splice site was not used in branching, is not connected to d5 in any obvious way and must compete with another d6 that is tethered to d5. It is possible that this round of splicing does not involve the d6 that is adjacent to 3'E^L at all and depends chiefly on the weak interactions known to be involved in 3' splice site selection.

The efficient use of the 3' splice site when the 3'E is in *cis* to d5 may explain, in part, why naturally occurring *trans*-spliced introns are arranged so that d5, d6 and the 3'E are on the same transcript. We have previously reported that the 5' splice site of a precursor containing 5'E plus intron domains 1-5 is efficiently cleaved and that d6-3'E RNA does not complement branching or splicing (3). Michel et al. have suggested that one reason for the arrangement of the naturally *trans*-splicing group II introns is to prevent premature cleavage at the 5' splice site (16). We suggest that having d5, d6 and 3'E on the same molecule meets the needs of both the first and second steps of splicing and, thereby, promotes efficient coupling between the two half reactions.

The current secondary structure model for group II introns proposes no specific contacts between d5 and d6 or between d5 and the 3' splice junction (2). Domain 5 of α 5 γ is tethered to d4 and d6 by short joining segments of 2 and 3 nt, respectively, that are not highly conserved. Thus, the joining segments are probably insufficient by themselves to orient precisely any one domain with respect to the rest of the folded structure. More likely, each domain makes a number of contacts with one or more of the other domains of the intron. Perhaps the helical regions of d5 and d6 adopt a coaxially stacked configuration. Since d5 and d6 are separated by only three nucleotides in this intron, stacking of the two helices seems reasonable by analogy with tRNA, where adjacent helices separated by just a few unpaired residues prefer to form a stack (17). Recently, several groups have proposed such stacking of adjacent helices in self-splicing group I introns (16, 18, 19). Stacking would allow the highly charged phosphates of those helices to be separated as much as possible.

Alternatively, d5 and d6 could interact side-by-side with the helical stems connected by the joining segment between them. Electrostatic repulsion would interfere with close packing unless sufficient counterions were present to shield the phosphates. In this arrangement, formation of hydrogen bonds between 2'OH groups on one helix and carbonyl groups or ring nitrogens on the other helix could stabilize side-by-side packing, as seen, for example, in the crystal structure of a small RNA helix (20); such interactions of helices or even stacked helices may exist in group I introns (18, 19). We would expect this sort of interaction to be stabilized by high salt concentrations; it may be significant that d5 provided in *trans* recruits d6 for branch formation only in the presence of very high salt concentrations. Further research will be required to distinguish between these two models for d5 interaction with d6.

Whatever the mode of d5 packing with d6, there clearly must be some interaction in addition to the covalent attachment through the joining segment. We have earlier shown that d5 interacts with d1 in a manner that is essential for at least the first splicing step. Most likely, both interactions occur without extensive canonical basepair formation, since d5 is composed largely of paired residues. Group II intron domains evidently fold to adopt locally favorable secondary structures that may then be able to interact with each other as folded entities, much like the subunits of a multisubunit protein enzyme. A recent study has proposed that U2 and U6 snRNAs interact with each other and with the branch acceptor region of the pre-mRNA to form an RNA structure that resembles d5 and d6 of group II introns (21). Since the two proposed helices are separated by a joining segment of 2 nucleotides it is possible that they may interact in one of the two ways that we have proposed here for d5 and d6.

ACKNOWLEDGEMENTS

This research was supported in part by research grant GM31480 from the National Institutes of Health and grant I-1211 from the Robert A. Welch Foundation to PSP, and by grant MCB-9206525 from the National Science Foundation to CLP. S.C.Boulanger was a predoctoral trainee supported by grant 5-T32-HL07360. We thank R.Padgett for a gift of HeLa cell S100 extract and for helpful comments on the manuscript. We appreciate the expert technical assistance of M.Zhang.

REFERENCES

1. Perlman, P.S., Peebles, C.L., and Daniels, C. (1990) in *Intervening Sequences in Evolution and Development* (Stone, E.M. and Schwartz, R.J., eds.) pp. 112–161. Oxford University Press, New York.
2. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene* **82**, 5–30.
3. Jarrell, K.A., Dietrich, R.C., and Perlman, P.S. (1988) *Molec. Cell. Biol.* **8**, 2361–2366.
4. Koch, J.L., Boulanger, S.C., Dib-Hajj, S.D., Hebbar, S.K., and Perlman, P.S. (1992) *Molec. Cell. Biol.* **12**, 1950–1958.
5. Schmelzer, C. and Muller, M.W. (1987) *Cell* **51**, 753–762.
6. van der Veen, R., Kwakman, J.H.J.M., and Grivell, L.A. (1987) *EMBO J.* **6**, 1079–1084.
7. Jacquier, A. and Michel, F. (1990) *J. Mol. Biol.* **213**, 437–447.
8. Wallasch, C., Morl, M., Niemer, I., and Schmelzer, C. (1991) *Nucleic Acids Res.* **19**, 3307–3314.
9. Jacquier, A. and Jacquesson-Breuleux, N. (1991) *J. Mol. Biol.* **219**, 415–428.
10. Ruskin, B. and Green, M. (1990) *Methods in Enzymol.* **181**, 180–188.
11. Arenas, J. and Hurwitz, J. (1987) *J. Biol. Chem.* **262**, 4274–4279.
12. Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L., and Perlman, P.S. (1988) *J. Biol. Chem.* **263**, 3432–3439.
13. Jarrell, K.A. (1988) *Cis and trans reactions of a self-splicing group II intron*. Ph. D. dissertation, Ohio State University.
14. Franzen, J. S., Zhang, M., and Peebles, C. L. (1993) *Nucleic Acids Res.* **21**, 627–634.
15. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H.-L. (1986) *Cell* **44**, 213–223.
16. Michel, F., Jaeger, L., Westhof, E., Kuras, R., Tihy, F., Xu, M.-Q. and Shub, S. A. (1992) *Genes & Devel.* **6**, 1373–1385.
17. Kim, S.-H. (1979) in *Transfer RNA: Structure, properties, and recognition* (Schimmel, P.R., Soll, D. and Abelson, J.N., eds.) pp. 83–100. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
18. Kim, S.-H. and Cech, T. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8788–8792.
19. Michel, F. and Westhof, E. (1990) *J. Mol. Biol.* **216**, 585–610.
20. Dock-Bregeon, A.C., Chevier, B., Podjarny, A., Johnson, J., de Bear, J.S., Gough, G.R., Gilham, P.T., and Moras, D. (1989) *J. Mol. Biol.* **209**, 459–474.
21. Madhani, H. D. and Guthrie, C. (1992) *Cell* **71**, 803–817.