Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 10658–10662, November 1993
Biochemistry

The phylogenetically predicted base-pairing interaction between α and α′ is required for group II splicing *in vitro*

CAROLE L. HARRIS-KERR*, MINCHENG ZHANG, AND CRAIG L. PEEBLES†

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Communicated by Mary Edmonds, August 20, 1993

ABSTRACT **The correct folding of group II introns apparently depends on multiple tertiary base-pairing interactions. Understanding the relationship between spliceosome and group II splicing systems ultimately requires a three-dimensional model for both structures. In turn, successful modeling depends at least in part on identifying tertiary base pairings. Sequence elements α and α′ are partners in a potential interaction of ≈6 base pairs that can be identified within domain 1 of most group II introns. In comparisons between related introns, α and α′ maintain their potential for Watson–Crick base pairing, even though their primary sequences can vary [Michel, F., Umesono, K. & Ozeki, H. (1989) *Gene* 82, 5–30]. Substitutions were constructed at α and α′ for a block of 6 bases each in the group II intron a5γ, the last intron of the *COXI* gene from the mitochondrial DNA of *Saccharomyces cerevisiae*. Each substitution was defective for self-splicing, while the compensatory double derivative was restored to active splicing. The α–α′ interaction is required for the first step of splicing—that is, recognition of the 5′ splice junction and transesterification with the branch site—since the derivative transcripts displayed little or no activity. The compensatory double derivative produced lariat introns and spliced exons with normal structures, showing that splicing activity and precise recognition were restored. We conclude that the α–α′ base pairing is necessary for efficient self-splicing by intron a5γ under several conditions. This result also provides an additional constraint for any three-dimensional model of group II intron structure.**

Group II introns are generally considered to be a relevant ribozyme model system for the eukaryotic spliceosome, a complex containing several small nuclear RNAs (snRNAs) and numerous protein factors (reviewed in ref. 1). A number of parallels have been identified between pre-mRNA splicing by the eukaryotic spliceosome and self-splicing by group II introns (1–4). Assembly of the spliceosome involves a number of different base-pairing interactions between the snRNAs and with the pre-mRNA (1, 2, 4). For example, U1 snRNA base pairs to the 5′ splice junction (5′SJ), U2 snRNA base pairs with the branch site, and U6 snRNA base pairs with U2 snRNA (1). This latter interaction is predicted to resemble domain 5 of group II introns (5). Correct folding of self-splicing introns depends on tertiary base-pairing interactions (6–8). These contacts are necessary to achieve active structures and to stabilize the intermediates for both group I (7, 8) and group II (6, 8) introns. Such interactions also provide valuable connections for building a structural model (for example, see ref. 9).

Constructing a three-dimensional structure model for group II introns is one of our long-term goals. This will aid in understanding their splicing function and the relationship between group II and spliceosome reaction mechanisms. Experimental results must establish connections between different parts of the intron in order to develop a valid three-dimensional model. Conserved, functionally relevant tertiary base-pairing contacts offer particularly important constraints, since they bring together two distant segments of the linear primary sequence in the active, folded form of the intron. Thus, we felt that it was important to establish the functional relevance of the predicted tertiary base-pairing interactions within group II introns. In this report, we show that base pairing occurs between the sequence elements α and α′ of group II introns.

Several other tertiary base-pairing interactions are known within group II introns. Considering just domain 1 and the adjacent 5′SJ, these include EBS1–IBS1, EBS2–IBS2, and ε–ε′ (refs. 10 and 11; see Fig. 1). The tertiary base-pairing interaction designated α–α′ was predicted to involve two sites within domain 1 (6, 10). These are α, the terminal loop of subdomain B, and α′, a bulging internal loop on the 3′ side of the helix of subdomain D3 (see Fig. 1; ref. 10). Often α–α′ consists of 6 base pairs, although both longer and shorter base-pairing regions are recognized (6). In naturally occurring introns, compensating primary sequence changes are known that maintain base pairing between α and α′. For example, such substitutions are evident between the structurally similar introns a5γ and b1 from the mitochondrial DNA of *Saccharomyces cerevisiae* or between the homologous ORF167/1 introns of *Marchantia polymorpha* and *Nicotiana tabacum* chloroplasts (6). In that compilation, α and α′ were identified for nearly all introns of subgroup IIA and for many of subgroup IIB (6). However, potential base-pairing elements corresponding to α and α′ have not yet been identified for all of the group II introns, suggesting that the α–α′ interaction may be lacking from some members of the group. Such a pattern could indicate that α–α′ is an optional feature with a limited role in splicing or that α–α′ merely represents a potential base pairing that is not realized structurally.

Although phylogenetic covariation suggests that base pairing between α and α′ is conserved in evolution and therefore functionally important, a more direct analysis could establish that α–α′ base pairing is necessary for self-splicing and reveal which splicing reactions are affected by loss of the α–α′ interaction. Since neither α nor α′ is predicted to contact the 5′SJ, the branch site, or the 3′SJ, it was not obvious how a defect in the α–α′ interaction might be manifested. To assess the role of α and α′, we introduced a 6-base substitution at α, another 6-base substitution at α′, and then combined these substitutions to restore complementarity. These variants were constructed within intron a5γ, the last intron of the *COXI* gene from yeast (12). We then tested the ability of these derivatives to perform self-splicing under several conditions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: snRNA, small nuclear RNA; 5′SJ, 5′ splice junction; 3′SJ, 3′ splice junction.

*Present address: Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, VA 23298.

†To whom reprint requests should be addressed at: 365 Crawford Hall, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260.

Our results establish that α–α′ base pairing is necessary for efficient self-splicing of intron a5γ. Furthermore, the α–α′ interaction is required for the first step of splicing, recognition of the 5′SJ, and transesterification to the branch site.

MATERIALS AND METHODS

Plasmids and Constructions. Plasmid pJD20 (13) represents the parental wild-type construct for all derivatives of the self-splicing intron a5γ described here. Plasmid pJD20 yields a 1522-nt transcript when cut with *Hin*dIII and transcribed by T7 RNA polymerase. The resulting spliced exons are 635 nt; the excised intron is 887 nt; the 5′ exon is 296 nt; and the 3′ exon is 339 nt. Plasmid DNAs were isolated by the alkaline lysis method from cultures grown overnight in LB-Amp (Luria–Bertani medium plus ampicillin at 100 μg/ml; ref. 14).

Mutagenic Oligonucleotides and Site-Directed Mutagenesis. All oligonucleotides were made at the University of Pittsburgh DNA Synthesis Facility on the Applied Biosystems model 380A instrument. Oligonucleotides were purified by polyacrylamide gel electrophoresis; sequence coordinates are specified relative to the long form of the *COXI* gene (12). Mutagenesis was performed essentially according to the method of Kunkel (15) using bacterial and helper phage strains obtained from Bio-Rad. Plasmid pJD20 was prepared as single-stranded uracil-containing DNA after transformation of *Escherichia coli* CJ236 (*dut*⁻, *ung*⁻) by growth at 37°C in the presence of ampicillin, infection with the helper phage M13-K07, and after selection by adding chloramphenicol at 30 μg/ml.

The mutagenic oligonucleotide for the α′ substitution was CK90-221 (5′-CAAATTATATTACTAATGGTTATATGAGGGCGAGTATGCTCATTTCCAATAC-3′), which is complementary to positions 8943–8994, except for the underlined change. The mutagenic oligonucleotide for the α substitution was CK90-222 (5′-CACTTGTAATTAGACTTATATTTCGCCCAATATAATTTATCCATTATTATATGG-3′), which is complementary to positions 8653–8706, except for the underlined change. Mutagenesis mixtures were transformed into *E. coli* DH5αF′ by the method of Hanahan (16). Five isolates of each were retained for plasmids pCK21 (α′ substitution) and pCK22 (α substitution). The double-derivative plasmid pCK2122 (both α and α′ substitutions) was made by transforming plasmid pCK21 into CJ236, isolating single-stranded uracil-containing DNA, annealing primer CK90-222, synthesizing the second strand, and transforming DH5αF′ as described. Recovery of the designed changes at α and α′ was confirmed by DNA sequencing using Sequenase (United States Biochemical) with primers CP4 (5′-CAGTCAAAGTTCCTTCATC-3′, complementary to positions 8791–8809) and CP91-1024 (5′-GGGTTTATTCTGTTTTATC-3′, complementary to positions 9024–9042). Plasmid pCK2122 was sequenced through both α and α′ sites.

Transcription and Self-Splicing Reactions. Radioactively labeled RNA was made by transcription of *Hin*dIII-cut plasmid DNA with T7 RNA polymerase in the presence of 50 μCi of [α-^{32}P]UTP (650 Ci/mmol; 1 Ci = 37 GBq; ICN) in 50-μl reaction mixtures with 0.8 mM each of the four unlabeled ribonucleotides (Pharmacia). Reaction mixtures were incubated for 1 h at 37°C and extracted with 1 vol of phenol/chloroform/isoamyl alcohol (25:24:1). The aqueous phase was centrifuged through columns of Sephadex G-50, precipitated with ethanol, and dissolved in 10 mM Tris·HCl, pH 7.5/1 mM Na_3EDTA (TE) plus 0.1% SDS. The yield and final concentration of transcripts were estimated by incorporation of radioactivity.

For splicing, transcripts were incubated for various times in 10 μl of one of the following splicing buffers: buffer AS, 0.5 M $(NH_4)_2SO_4$/100 mM $MgCl_2$/40 mM Tris·HCl, pH 7.5/0.1% SDS; buffer K, 0.5 M KCl/100 mM $MgCl_2$/40 mM Tris·HCl, pH 7.5/0.1% SDS; buffer M, 100 mM $MgCl_2$/40 mM Tris·HCl, pH 7.5/0.1% SDS (13). Reactions were stopped by adding loading dye, and the products were analyzed by electrophoresis through denaturing 4% polyacrylamide gels before autoradiography.

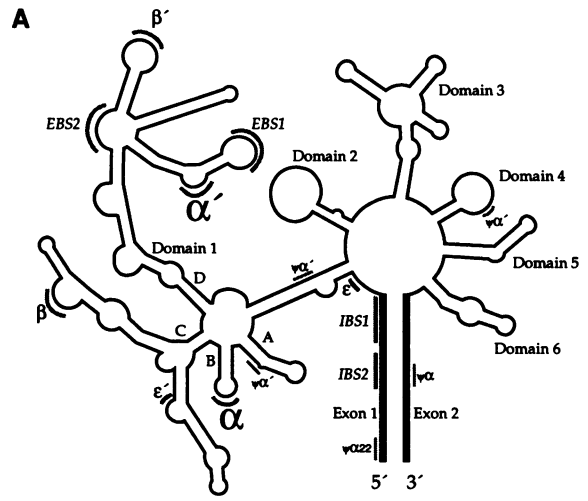
Product Purification and Primer Extension. Splicing products were isolated from 100-μl reaction mixtures after electrophoresis on 30 cm, 4% polyacrylamide gels for 4 h at 700 V. Individual products were located after autoradiography and excised with a new razor blade. Gel slices were crushed and shaken at 37°C for at least 4 h in an elution buffer consisting of 0.3 M $NaO_2C_2H_3$ (pH 5), 10 mM Na_3EDTA, and 1% phenol. The aqueous portion was recovered by filtration. Splicing products were then precipitated three times with ethanol, dissolved in water, and used as templates for reverse transcription with deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate phosphorothioate (14). Primer CP4 yields a product of 190 nt with wild-type lariat intron as the template (17). The labeled cDNAs were treated with RNase A, precipitated with ethanol, and suspended in an equal volume of formamide loading dye. The products were separated on 6% polyacrylamide gels in parallel with dideoxynucleotide sequencing ladders made from the same primer on the wild-type template.

PCR and Sequencing of Spliced Exons. Purified RNAs were converted to unlabeled cDNAs by reverse transcription from primer CP90-305 (5′-GAGGACTTCAATAGTAGTATCCTGC-3′, complementary to positions 9545–9569). The cDNAs were treated with RNase A and amplified by a PCR (94°C, 1 min; 50°C, 2 min; 72°C, 1 min; 30 cycles). The primers were CP90-305 and CK90-202 (5′-CACAATGGGTGGTTTAACTGG-3′, corresponding to positions 8531–8551). PCR products were separated on agarose gels, captured on a DEAE membrane (NA-45; Schleicher & Schuell), eluted with high-salt 1 M NaCl/10 mM Tris·HCl, pH 7.5/1 mM Na_3EDTA (NET) buffer, and precipitated with ethanol. The purified PCR products were ligated to pUC119 that had been cleaved with *Sma* I. The ligation mixtures were transformed into *E. coli* JM101 and plated on LB with carbenicillin (50 μg/ml) plus 5-bromo-4-chloro-3-indolyl β-D-galactoside and isopropyl β-D-thiogalactopyranoside to identify those with inserts as white colonies. Plasmids were recovered and their cDNA inserts were sequenced.

RESULTS

Design and Construction of α and α′ Derivatives. Substitutions were introduced by site-directed mutagenesis (15) at α and α′ within domain 1 of the group II intron a5γ. These sites were predicted to interact by base pairing (ref. 10; Fig. 1*A*). The wild-type α′ sequence 5′-GAUAAA-3′ was changed to 5′-UCGCCC-3′ in the single substitution, a5γ–α'_{21} (Fig. 1*B*). A sequence hexamer identical to wild-type α′ occurs at three other sites within the intron; such pseudo-α′ (ψα′) sites are not phylogenetically conserved (Fig. 1*A*). At least two of these ψα′ sequences are involved in secondary structure base pairing and are thus not available to interact with α. The sequence hexamer used as the α'_{21} substitution does not occur anywhere else in the self-splicing transcript.

The wild-type α sequence 5′-UUUAUC-3′ was changed to 5′-GGGCGA-3′ in the other single substitution, a5γ–α_{22} (Fig. 1*B*). One pseudo-α (ψα) sequence occurs in the 3′ exon of the self-splicing transcript; one pseudo-α_{22} ($\psi\alpha_{22}$) sequence occurs in the 5′ leader of the transcript derived from the polylinker (Fig. 1*A*). The compensatory double derivative a5γ–α'_{21}/α_{22} contains both α′ and α substitutions and was made from the α′ variant, a5γ–α_{22} (Fig. 1*B*). All changes were verified by DNA sequencing of the final plasmids.



B

α_{wt} 5′-auuUUUAUCaau-3′ α_{wt} 5′-auuUUUAUCaau-3′
α_{22} 5′-auuGGGCGAaau-3′ α'_{wt} 3′-acuAAAUAGcau-5′
α'_{wt} 5′-uacGAUAAAuca-3′ α_{22} 5′-auuGGGCGAaau-3′
α'_{21} 5′-uacUCGCCCuca-3′ α'_{21} 3′-acuCCCGCUcau-5′

FIG. 1. Secondary structure of group II introns, emphasizing α and α' sequence elements. (*A*) Transcript containing the self-splicing group II intron a5γ. Intron (thin line) is depicted as a series of secondary structure domains (Domains 1–6) bounded by base-paired helical segments (paired straight lines). Domain 1 is shown as four subdomains (A–D). No attempt is made to diagram the folding pattern inside domain 2 or 4. Thick straight segments mark exons 1 and 2. Several tertiary base-pairing interactions are emphasized by shaded overbars: *IBS2* pairs with *EBS2*, *IBS1* pairs with *EBS1*, ε pairs with ε', α pairs with α' (as shown here); it has been proposed that β pairs with β'. Phylogenetic and other evidence defining these interactions is mentioned in the text and reviewed elsewhere (6). Sites with hexamer sequences identical to α' are tagged as $\psi\alpha'$ with light overbars. These occur at coordinates 8666–8671 in subdomain A of domain 1, 9024–9029 in the helical stem of domain 1, and 9380–9385 in a portion of domain 4 lacking a conserved secondary structure. Hexamer sequence identical to α occurs at 9776–9781 ($\psi\alpha$ in exon 2). The lone hexamer sequence match to a substitution is similarly marked as $\psi\alpha_{22}$ in the 5′ polylinker part. (*B*) (*Left*) Primary sequences of α and α' for wild type (α_{wt} and α'_{wt}), a5γ-α'_{21} (α'_{21}), and a5γ-α_{22} (α_{22}). (*Right*) Base-pairing interactions for wild type and double-derivative a5γ-α'_{21}/α_{22}. Sequences are listed 5′ to 3′ except in the base-pairing diagram, where italics emphasize sequences listed 3′ to 5′. The α and α' elements are shown in capital letters within the longer sequence context shown.

The replacement sequences for α and α' were made as 6-base transversions. This design was chosen because many α segments are pyrimidine-rich, and the complementary α' is correspondingly purine-rich. Other α sites contain RYRY or RYY motifs (where R is purine and Y is pyrimidine) (6). Although the significance of these features is not known, we reasoned that our design would reveal an effect if this sequence bias was important for function. To avoid simply exchanging the sequences of α and α', we applied the following base substitution rule: U to G, A to C, G to U, and C to A. The resulting double derivative has a purine-rich α and a pyrimidine-rich α'. This double-derivative transcript offers fewer alternative interactions for α and α' than occur in the wild type. We supposed that these 6-base substitutions would provide sufficiently drastic effects on the α–α' interaction to reveal the mutant phenotypes clearly, while the double derivative a5γ-α'_{21}/α_{22} would regain complementarity between the predicted α and α' sites.

Splicing Activity of the α and α' Variants in $(NH_4)_2SO_4$. Initially, each transcript derivative was incubated in splicing buffer containing 0.5 M $(NH_4)_2SO_4$ (13). The a5γ-α'_{21} derivative was inactive (Fig. 2*A Left*). By contrast, most of the wild-type transcript had reacted by 5 min (Fig. 2*A Right*). The a5γ-α_{22} variant showed a little activity and produced some lariat intron plus a small amount of spliced exons by 60 min (Fig. 2*B*). The a5γ-α_{22} variant reacts more slowly than wild type, although it is significantly more reactive than a5γ-α'_{21}. The a5γ-α_{22} variant also showed a sharply reduced extent of reaction after 60 min, estimated as <10% of wild type. Thus, it is quite evident that both single substitutions are seriously defective for splicing.

The double derivative a5γ-α'_{21}/α_{22} restores the potential for base pairing between α and α' (Fig. 1*B*) and significantly restores splicing activity. The double derivative produces the wild-type distribution of products in $(NH_4)_2SO_4$ splicing buffer (Fig. 2*C*). However, the extent of the reaction for the double derivative remained somewhat less than that for wild type. Nevertheless, it is clear that restoring base pairing to the α–α' interaction substantially restored function with the double replacement.

Splicing Activity of the α and α' Variants in KCl. Parallel experiments were performed in splicing buffer containing 0.5 M KCl (ref. 13; Table 1). The KCl reaction condition is often regarded as permissive for mutant constructs (18). The a5γ-α'_{21} substitution remained completely unreactive. With the a5γ-α_{22} substitution, barely detectable lariat and linear intron signals were produced by 20 min. In comparison, most wild-type transcripts reacted by 10 min. The splicing activity of the a5γ-α_{22} variant was significantly less in splicing buffer with KCl than with $(NH_4)_2SO_4$. The activity of the double-derivative a5γ-α'_{21}/α_{22} was restored nearly to the wild-type rate and extent and with a normal distribution of products in splicing buffer containing KCl. It is important to point out that the wild-type transcript yields a product profile in KCl different than in $(NH_4)_2SO_4$. Separate 5′ and 3′ exons accumulate in KCl, and these are derived from hydrolysis (spliced exon reopening) catalyzed by the intron (13). In KCl, the products of the double-derivative a5γ-α'_{21}/α_{22} showed somewhat lower levels of separate exons than the wild type.

Splicing Activity of the α and α' Variants in 0.1 M $MgCl_2$. Several group II intron derivatives have little or no splicing activity in low salt, but many such mutants display at least partial activity with added high salt, such as 0.5 M $(NH_4)_2SO_4$ or 0.5 M KCl (for example, see ref. 18). Thus, the 0.1 M $MgCl_2$ splicing buffer is generally used as a restrictive condition when testing mutants. To verify that the α and α' substitutions were defective and that the double derivative was restored for activity, we tested these transcripts in splicing buffer containing 0.1 M $MgCl_2$ but no added salt (Table 1). Both of the single substitutions, a5γ-α'_{21} and a5γ-α_{22}, were completely inactive in this 0.1 M $MgCl_2$ splicing buffer. The double-derivative a5γ-α'_{21}/α_{22} accumulated the wild-type pattern of products, although it reacted more slowly and to a reduced extent when compared to wild type. Thus, the 0.1 M $MgCl_2$ splicing buffer is indeed the most restrictive condition for these derivatives. This reaction condition yields the simple response that activity is lost for both single substitutions and restored with the compensatory double derivative.

Characterization of Splicing Products from the Double-Derivative a5γ-α'_{21}/α_{22}. To determine whether the splicing products from the double replacement had the same structure as wild type, spliced exons and lariat introns were recovered from splicing reactions done with 0.5 M $(NH_4)_2SO_4$. The lariat introns were analyzed by primer extension to map their 5′ ends (Fig. 3). The cDNAs generated by primer extension were precisely the same size for both a5γ-α'_{21}/α_{22} and wild-type lariats. This indicates that the a5γ-α'_{21}/α_{22} transcript accurately selects the 5′SJ and performs transesterification at the same position as wild type. This was confirmed

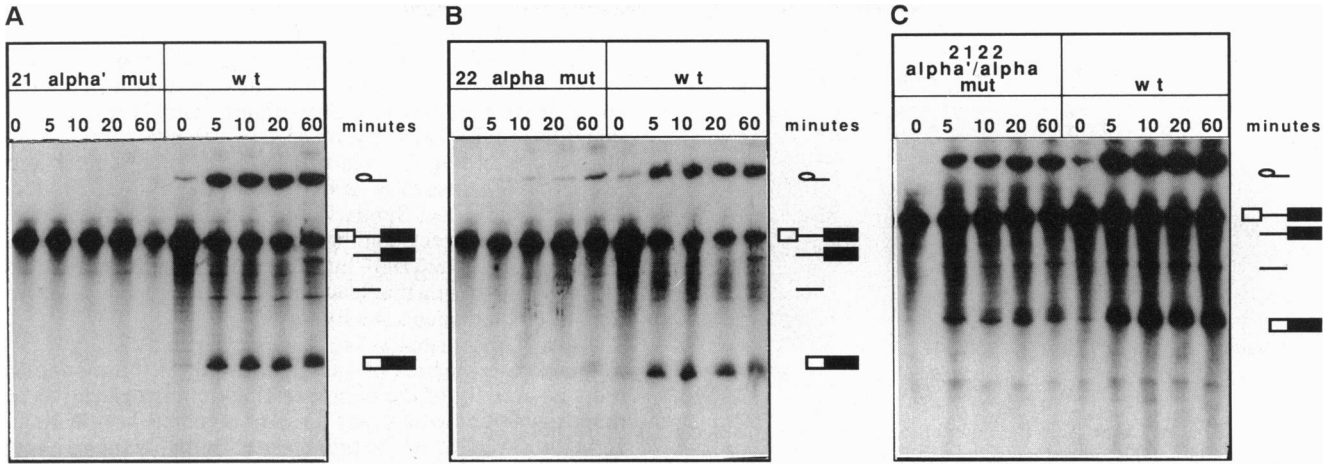


FIG. 2. Splicing activity by α and α' variants in $(NH_4)_2SO_4$. Transcripts were incubated in splicing buffer AS at 45°C for the times marked above each lane. (*A*) Wild type (wt) and a5γ–α'_{21} (21 alpha' mut). (*B*) Wild type (wt) and a5γ–α_{22} (22 alpha mut). (*C*) Wild type (wt) and a5γ–α'_{21}/α_{22} (21/22 alpha'/alpha mut). Reactions were stopped by adding loading dye, products were separated on denaturing 4% polyacrylamide gels for 4 h at 700 V, and gels were exposed to XAR film for autoradiography. Thin line, intron; open box, exon 1; solid box, exon 2; tailed-circle, intron lariat. Symbols are not drawn to scale. The wild type, but not the double derivative, undergoes a little splicing during the transcription reaction, accounting for the wild-type lariat visible in the 0-min lane.

by performing a similar experiment on unfractionated splicing products and obtaining the same result (data not shown). To determine the structure of the spliced exons from the double-derivative a5γ–α'_{21}/α_{22} transcript, we performed reverse transcription, PCR amplification, cloning, and DNA sequencing (Fig. 4). Six isolates were tested, and all were identical to wild type, containing the sequence 5′-CATTTTCACTATG-3′. In summary, the lariat intron 5′ ends are identical, and the sequences of the spliced exon cDNAs from a5γ–α'_{21}/α_{22} and wild type are the same. Together, these results demonstrate that the compensatory double derivative actually restores normal splice site choice for both steps of the splicing reaction.

DISCUSSION

Successful use of a purine-rich replacement for the usually pyrimidine-rich α' sequence element (and the reciprocal substitution for α) suggests that the primary sequence or base composition of these elements is much less important than the success of their mutual base-pairing interaction. The products of the compensatory double-derivative a5γ–α'_{21}/α_{22} are lariat introns and spliced exons identical to wild type. Clearly the most important role of the α and α' sites in the self-splicing intron a5γ is fulfilled by Watson–Crick base pairing.

Table 1. Relative splicing activity of α and α' variants

Transcript	Splicing buffer	Activity level	Product pattern
a5γ–α_{22}	AS	Reduced	Normal
a5γ–α'_{21}	AS	None	None
a5γ–α'_{21}/α_{22}	AS	Normal	Normal
a5γ–α_{22}	K	Trace*	Normal*
a5γ–α'_{21}	K	None	None
a5γ–α'_{21}/α_{22}	K	Normal	Normal
a5γ–α_{22}	M	None	None
a5γ–α'_{21}	M	None	None
a5γ–α'_{21}/α_{22}	M	Reduced	Normal

Normal, >50% of wild-type activity; Reduced, 10–50% of wild-type activity; Trace, <10% of wild-type activity.
*In some experiments, faint autoradiographic signals could be seen at later reaction times for some of the positions corresponding to normal products (for example, intron lariat), but some of the other products could not be detected (for example, spliced exons).

Results from several conditions show that both single substitutions, a5γ–α'_{21} and a5γ–α_{22}, are inactive or have greatly reduced activity (Fig. 1; Table 1). The failure to initiate splicing that is the result of making substitutions at either α or α' establishes that this interaction plays an important role in the first step of the reaction. The return of accurate splicing in the compensatory double derivative verifies that α and α' make contact by base pairing and that this is necessary for splicing to occur. Since the α–α' base-pairing interaction is a tertiary contact, one reasonable interpretation is that both single substitutions fail to fold correctly, leading to failure of 5′SJ selection or transesterification activity. The observation that a5γ–α_{22} splices to a limited extent under at least one condition (Fig. 2) suggests that intron folding is impaired but not wholly disrupted. Apparently, a minority population of transcripts fold correctly and react, while the remainder do not initiate the

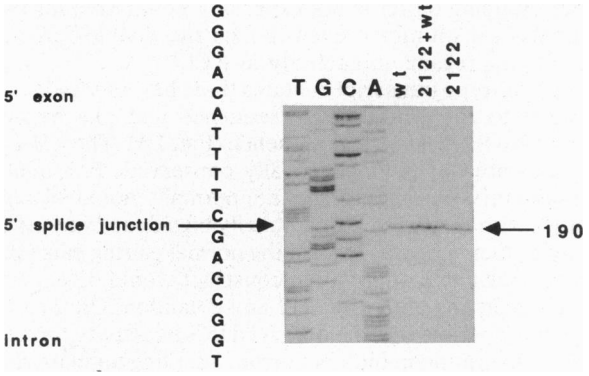


FIG. 3. Intron lariats from the a5γ–α'_{21}/α_{22} transcript are identical to wild type. Intron lariats from the a5γ–α'_{21}/α_{22} transcript (lane 2122) or from wild type (lane wt) were analyzed by primer extension. The center lane (lane 2122+wt) contains a mixture of equal amounts from the samples loaded in the two adjacent lanes. The primer was CP4, a 19-mer that hybridizes within domain 1 and yields a cDNA product from the wild-type lariat of 190 nt (ref. 17; arrow). Sequencing ladder was derived from primer CP4 on a wild-type template and appears as 4 lanes (T, G, C, and A). Part of the DNA sequence shown on the gel is listed 5′ to 3′ (top to bottom), and distinct regions of the transcript are designated on the far left.

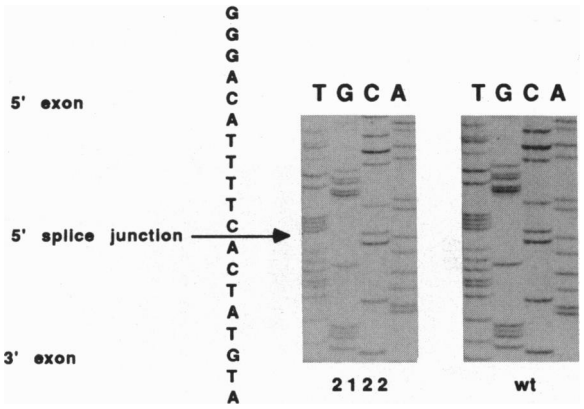


FIG. 4. Spliced exons from the compensating double-derivative a5γ–α'_{21}/α_{22} have the wild-type sequence. Cloned cDNAs were prepared from either wild type (wt) or a5γ–α'_{21}/α_{22} spliced exons (2122) by reverse transcription using CP90-305, a primer that anneals to the 3′ exon portion of the spliced exons. cDNAs were treated with RNase A and amplified by PCR using CK90-202 (5′ exon) and CP90-305 primers. PCR products were purified and cloned into pUC119. Plasmids from individual isolates were purified and sequenced. Sequencing reaction mixtures were separated on a 6% polyacrylamide gel, and the autoradiogram is shown (lanes T, G, C, and A). Part of the corresponding DNA sequence is listed 5′ to 3′ (top to bottom), and the regions of the spliced exons are noted on the far left.

reaction even after extended incubation. This outcome supports the suggestion by Michel *et al.* (6) that α–α′ base pairing might provide appropriate positioning of EBS1 and EBS2 relative to IBS1 and IBS2. Alternatively, the α–α′ base pairing might align elements of the catalytic center properly. In either case, splicing activity returns within the compensatory double-derivative a5γ–α'_{21}/α_{22}, although the splicing rate and extent remain somewhat less than those for wild type. Restored pairing does not fully compensate for altered primary sequences at α and α′. This suggests that α, α′, or the base-paired form of α–α′ might make additional sequence-dependent contacts within the functional intron.

Many group II intron derivatives are rescued to at least partial reactivity by using 0.5 M KCl in the splicing buffer (18). However, both single substitutions at α and α′ were essentially inactive under that condition (Table 1). Thus, 0.5 M KCl splicing buffer is not especially permissive for these particular substitutions, even though the double-derivative a5γ–α'_{21}/α_{22} regains full activity in KCl.

The wild-type transcript contains three hexamer sequences identical to the wild-type α′ sequence and one hexamer identical to the wild-type α sequence (Fig. 1*A*). These ψα and ψα′ sites are not phylogenetically conserved. Presumably, such potential pairing partners are normally not available to interact with α or α′ in a functionally significant way. If one of these other sites was instead the normal pairing partner for α or α′, our double-derivative construct would be expected not to regain any activity under any condition. On the other hand, it is possible that the observed loss of activity for either single substitution could result from adopting an inappropriately folded structure, rather than from loss of the α–α′ base pairing itself. For example, the wild-type α′ sequence is present and unpaired in the a5γ–α_{22} transcript; that α′ has the potential to pair inappropriately with the ψα sequence in the 3′ exon. Likewise, the new sequence introduced as α_{22} in the a5γ–α_{22} transcript has no complementary hexamer sequence available for base pairing, although partial matches are present. Therefore, both α′ and α_{22} lack normal pairing partners in the a5γ–α_{22} transcript and could pair inappropriately. Notably, we observed some activity with a5γ–α_{22} in $(NH_4)_2SO_4$ splicing buffer (Fig. 2), so at least some transcripts are folded correctly, and thus no alternative folding could be strongly favored. This argues against the conclusion that the reduced activity of a5γ–α_{22} results from a stable alternative folding for this transcript. It seems much more likely that it is loss of α–α′ that leads directly to the reduced activity of a5γ–α_{22}. Similarly, we cannot rule out the possibility that incorrect folding of the transcript leads to the inactivity of a5γ–α'_{21}/α_{22}, but we suggest that it is the absence of α–α′ base pairing that leads to the observed loss of activity for both of the single substitutions.

Altering the primary sequences of α and α′ appears to influence the kinetics of splicing, based on the moderately reduced activity of the compensating double derivative. We think that α and α′ are not directly involved in selecting or activating the 5′SJ or the branch site. Probably the α–α′ base pairing functions in intron folding, perhaps positioning the arms of domain 1 in the proper conformation for splicing. Those naturally occurring introns that lack an identifiable α–α′ interaction may accomplish the same end by other base-pairing interactions that have not yet been identified, or they may rely on auxiliary protein factors to fold correctly.

Despite an apparently indirect role in the splicing mechanism, the α–α′ interaction is clearly important, because the lack of α–α′ base pairing results in severely impaired self-splicing activity for both single substitutions in the a5γ intron. The α–α′ base-pairing interaction positions the loop at the end of subdomain B near the side of the D3 helix in domain 1, providing an important constraint for a future three-dimensional model of group II intron structure.

We thank our colleagues Mary Edmonds, James Franzen, Paula Grabowski, Phillip Perlman, and John Woolford for reading earlier versions of this manuscript and providing valuable criticism. This research was supported in part by grants from the American Cancer Society (NP-729) and the National Science Foundation (MCB-9206525) and by intramural support from the University of Pittsburgh.

1. Moore, M. J., Query, C. C. & Sharp, P. A. (1993) in *The RNA World*, eds. Gesteland, R. A. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 303–357.
2. Guthrie, C. (1991) *Science* **253,** 157–163.
3. Jacquier, A. (1990) *Trends Biochem. Sci.* **15,** 351–354.
4. Baserga, S. J. & Steitz, J. A. (1993) in *The RNA World*, eds. Gesteland, R. A. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 359–381.
5. Madhani, H. D. & Guthrie, C. (1992) *Cell* **72,** 803–817.
6. Michel, F., Umesono, K. & Ozeki, H. (1989) *Gene* **82,** 5–30.
7. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
8. Cech, T. R. (1993) in *The RNA World*, eds. Gesteland, R. A. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 239–269.
9. Michel, F. & Westhoff, E. (1990) *J. Mol. Biol.* **216,** 585–610.
10. Jacquier, A. & Michel, F. (1987) *Cell* **50,** 17–29.
11. Jacquier, A. & Michel, F. (1990) *J. Mol. Biol.* **213,** 437–447.
12. Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A. & Macino, G. (1980) *J. Biol. Chem.* **255,** 11927–11941.
13. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. (1988) *J. Biol. Chem.* **263,** 3432–3439.
14. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. M., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
15. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
16. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
17. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. & Cheng, H.-L. (1986) *Cell* **44,** 213–223.
18. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K. & Perlman, P. S. (1992) *Mol. Cell. Biol.* **12,** 1950–1958.