Volume 13 Number 6 1985 Nucleic Acids Research

Sequence requirements for self-splicing of the *Tetrahymena thermophila* pre-ribosomal RNA

James V.Price[1], Gary L.Kieft[2+], Jeffrey R.Kent[2§], Eric L.Sievers[2*] and Thomas R.Cech[1,2]

[1]Department of Molecular, Cellular and Developmental Biology, and [2]Department of Chemistry, University of Colorado, Boulder, CO 80309, USA

Received 18 January 1985; Accepted 14 February 1985

ABSTRACT

The sequence requirements for splicing of the Tetrahymena pre-rRNA have been examined by altering the rRNA gene to produce versions that contain insertions and deletions within the intervening sequence (IVS). The altered genes were transcribed and the RNA tested for self-splicing in vitro. A number of insertions (8-54 nucleotides) at three locations had no effect on self-splicing activity. Two of these insertions, located at a site 5 nucleotides preceding the 3'-end of the IVS, did not alter the choice of the 3' splice site. Thus the 3' splice site is not chosen by its distance from a fixed point within the IVS. Analysis of deletions constructed at two sites revealed two structures, a hairpin loop and a stem-loop, that are entirely dispensable for IVS excision in vitro. Three other regions were found to be necessary. The regions that are important for self-splicing are not restricted to the conserved sequence elements that define this class of intervening sequences. The requirement for structures within the IVS for pre-rRNA splicing is in sharp contrast to the very limited role of IVS structure in nuclear pre-mRNA splicing.

INTRODUCTION

The pre-ribosomal RNA of Tetrahymena thermophila contains an intervening sequence which is processed in an unusual manner: the IVS is accurately excised and the exons are efficiently ligated in the absence of proteins. Guanosine (or GTP) is required for this reaction and becomes covalently linked to the 5' end of the IVS by a normal 3'-5'phosphodiester bond (1,2). A partial precursor RNA, transcribed in vitro from plasmid pIVS11, also displays this behavior (3). Purified intervening sequence can undergo a further reaction without proteins. A guanosine residue at the 3' end of the IVS provides a 3' hydroxyl for an attack on the bond between residues 15 and 16 of the IVS. This reaction results in the transfer of the phosphodiester bond from residue 15 to the 3' terminal guanosine, forming a circular IVS and releasing a 15 nucleotide fragment (4,5). The IVS cyclization reaction and the splicing reaction are similar with respect to their mechanism; therefore the IVS, which must provide the active site for cyclization, could also provide the active site for the splicing reaction.

The Tetrahymena ribosomal intron has been found to contain several sequence elements which are conserved in the major group (Group I) of fungal mitochondrial mRNA and rRNA introns (6-9). Four sequence elements called A, B, 9L and 2, containing 10-12 nucleotides each, are highly conserved on the basis of sequence. The degree of sequence conservation over long evolutionary distances suggests a functional role. These four elements are thought to pair with each other to form short helices, A with B and 9L with 2 (10,11,7-9). The pairing of these sequence elements is supported by phylogenetic evidence. When a sequence variant is observed at one site, the potential for base-pairing of the two elements is maintained by a second change in the complementary site. There are two other elements, 9R and 9R', which are not conserved in sequence but are always found in the same position relative to the fourconserved elements and are capable of base-pairing with each other. The actual pairing of these two loci is supported by mutational analysis of a yeast mitochondrial intron. An intron with a mutation in either 9R or 9R' is defective in splicing, but a double mutation that restores sequence complementarity is spliced normally (12). The intervening sequences which contain these conserved regions are thought to represent a family with a common core structure. In addition to structural similarities, it is apparent that similar biochemical mechanisms are involved in the processing of this class of intervening sequences (13-15; H.F. Tabak, personal communication).

Several models have been proposed for the secondary structure of the T. thermophila rRNA IVS (7-9). These models share many common properties. The model proposed by Cech et al. (9), shown in Figure 1, was used as a framework for our experimental work. In order to examine the structural requirements for the activity of the T. thermophila IVS, a number of insertions and deletions were constructed at four locations in the IVS. Evidence is presented that insertions and small deletions have little effect on splicing activity at all four sites. Larger deletions, however, were observed to eliminate all detectable activity. We conclude that some structures far from the splice sites, other than the conserved elements, are required for splicing activity.

MATERIALS AND METHODS

Materials

T4 DNA ligase, T4 polynucleotide kinase, radioisotopes and pSP62-PL were purchased from New England Nuclear, reverse transcriptase from Life Sciences and SP6 RNA polymerase from Promega Biotec. All restriction enzymes, synthetic linkers and BAL 31 were purchased from New England Biolabs. All diges-

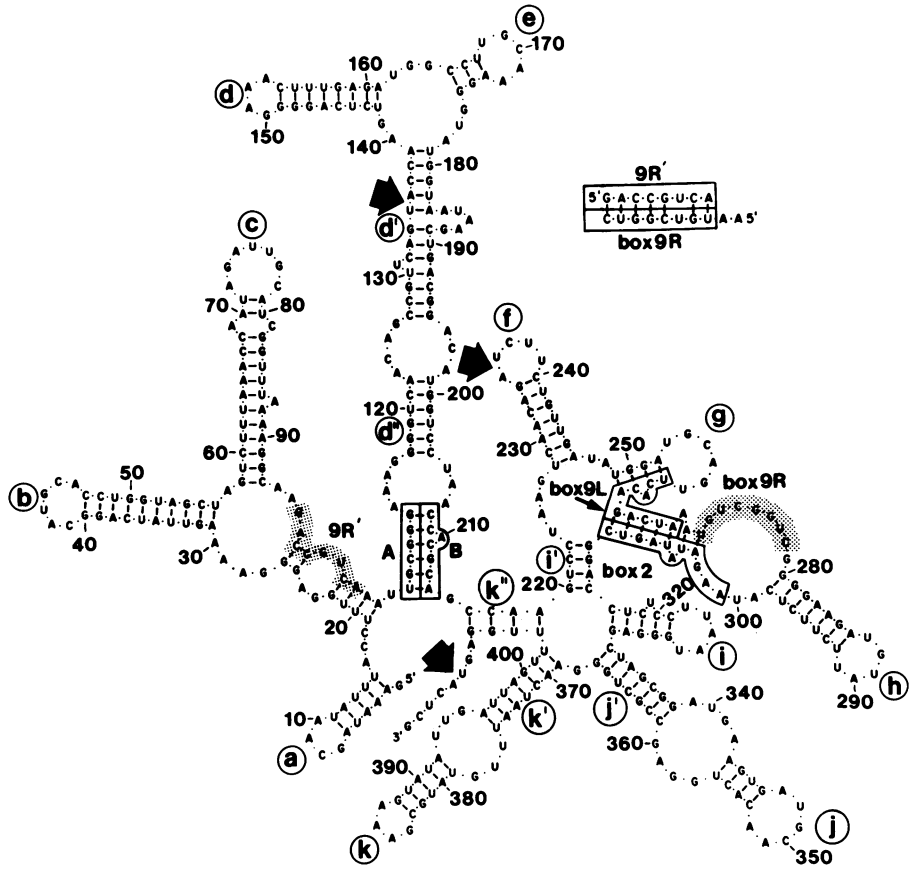

Figure 1. Model of the Secondary Structure of the T. thermophila IVS Showing Locations of Insertions Generated by Recombinant DNA Manipulations. One or more plasmids were constructed containing synthetic linkers inserted at each site indicated by an arrow. Six Pst I linkers (GCTGCAGC) were inserted at position 236 to give pIVS4. The other two positions were the sites of insertions of between one and five Bam HI linkers (CGGATCCG).

tions were carried out in buffers recommended by the supplier. E. coli RNA polymerase was supplied by Dr. Carol Cech. Synthetic oligonucleotide primers were obtained from Jeri Beltman and Dr. Marvin Caruthers.

Plasmid Constructions

All plasmids used in this study were derived from pIVS11 (3). pIVS11 contains a 1.6 kb Hind III fragment of the T. thermophila rDNA, including the 413 bp IVS, 261 bp of the 5' exon and 948 bp of the 3' exon. In pIVS11, this rDNA fragment can be transcribed from a *lac* UV5 promoter.

To construct pIVS4, pIVS11 was digested with restriction endonuclease Bgl II, treated with E. coli DNA polymerase I (large fragment) in the presence of all four dNTPs and religated with T4 DNA ligase in the presence of phosphorylated Pst I linkers. E. coli were transformed as described by Schleif and Wensink (16). Transformants were selected on LB plates containing 50 µg/ml ampicillin. Colonies were screened by mini-prep (as described by Holmes and Quigley (17), modified by addition of RNase A treatment, phenol/chloroform extraction and EtOH precipitation) and restriction analysis. A derivative of pIVS11 was isolated which contained six Pst I linkers at the Bgl II site.

Deletion libraries were constructed by digesting pIVS11 with either Bgl II (pGK series) or Sph I (pES series). 30 µg of plasmid were then digested with 4 units of nuclease BAL 31 at 30° C in 450 µl of 600 mM NaCl, 12 mM $CaCl_2$, 12 mM $MgCl_2$, 20 mM Tris (pH 8.0). At 1 min intervals 90 µl aliquots were transferred to tubes containing 44 µl 100 mM EDTA (pH 8.0). After ethanol precipitation, samples were resuspended in H_2O. One third of each sample was subjected to restriction analysis to determine the extent of BAL 31 digestion. A one min time point was selected for the deletion library at position 42 (approximately 4 to 30 base pairs deleted), and a three min time point was selected for the library at position 236 (approximately 15 to 50 base pairs deleted). These samples were treated with T4 DNA polymerase, in the presence of all four dNTPs, to increase the frequency of blunt ends, and religated with T4 DNA ligase. Each sample was re-cut with either Sph I (pES series) or Bgl II (pGK series) to eliminate transformation by molecules which had not been digested by BAL 31 at the restriction site. Transformation was carried out as above. Deletions were screened by mini-prep and restriction analysis.

Bam HI linker insertion derivatives of pIVS11 were constructed by partially digesting 112 µg of pIVS11 with 110 units of restriction endonuclease Rsa I for 4 min in a 300 µl reaction. The reaction was quenched by addition of 40 µl of 50 mM EDTA (pH 8.0), 1% SDS. Unit length linear plasmid was purified by gel electrophoresis on a 1% agarose gel followed by electroelution by the method of Dretzen et al. (18). After ethanol precipitation the fragment was suspended in water. The fragment was treated with T4 DNA ligase in the presence of a 50 fold molar excess of phosphorylated Bam HI linkers. Transformation was carried out as above. Transformants were screened by mini-prep and restriction analysis.

Synthesis of Pre-rRNA

Large scale plasmid preparations were done by the method of Holmes and Quigley (17). Transcription with E. coli RNA polymerase was done by the

method of Kruger et al. (3) using 10 mM NaCl instead of 5 mM $(NH_4)_2SO_4$. When splicing was desired during transcription, 50 mM NaCl was used. Transcription of pJK43-SP6, pJK15-SP6 and pJK29-SP6 was done with SP6 RNA polymerase in a solution containing 40 mM Tris (pH 7.5), 6mM $MgCl_2$, 10 mM DTT, 40mM spermidine, and .5 mM each ATP, CTP, GTP, UTP.

Splicing of Pre-rRNA

Splicing was carried out at 30°C in 30 mM Tris (pH 7.5), 100 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$. For splicing without end-labeling, 500 μM GTP was included. For splicing with end-labeling, transcription products were gel-purified or fractionated on Sephadex G-50-150 to remove unincorporated nucleotides. Splicing was then carried out in the presence of 2 μM α^{32}P-GTP unless otherwise indicated.

Gel Electrophoresis

Analytical and preparative gel electrophoresis of RNA was performed on 4% polyacrylamide gels containing 8 M urea. Samples were heated 5 min in urea-containing sample buffer before loading. Running buffer contained 89 mM Tris, 89 mM boric acid, 2.5 mM EDTA (pH 8.3). After soaking out the urea, gels were dried and autoradiographed or, in the case of tritium fluorography, soaked in 1 M sodium salicylate for 30 min, rinsed several times in fresh water and dried. Dried gels were exposed to Kodak XAR-5 x-ray film (pre-flashed in the case of fluorography) using a Dupont Cronex intensifying screen at -70° C. Films were developed for 5 min with Kodak D-19 developer and fixed with Kodak rapid fixer for 5 min at approximately 68° C. X-ray films were scanned and peaks integrated as described by Bass and Cech (2).

RNA was recovered from crushed polyacrylamide gel slices by soaking overnight in 3-5 volumes of 0.5 M $NH_4C_2H_3O_2$, 0.1mM EDTA (pH 7.5), 0.1% SDS at 22° C.

RNA Sequencing

Dideoxynucleotide sequencing was performed as described by Inoue and Cech (19) except dNTPs were used at a final concentration of 400 μM each and a ddNTP was added to a concentration of 80 μM in each sequencing reaction.

RESULTS

An Insertion at Position 236

Six synthetic Pst I linkers (GCTGCAGC) were inserted at a Bgl II site at position 236 of the IVS in pIVS11 to generate pIVS4. The methods used involved duplication of the 4 nucleotide overhang generated by Bgl II digestion, so a total of 52 base pairs were inserted. RNA transcribed in vitro

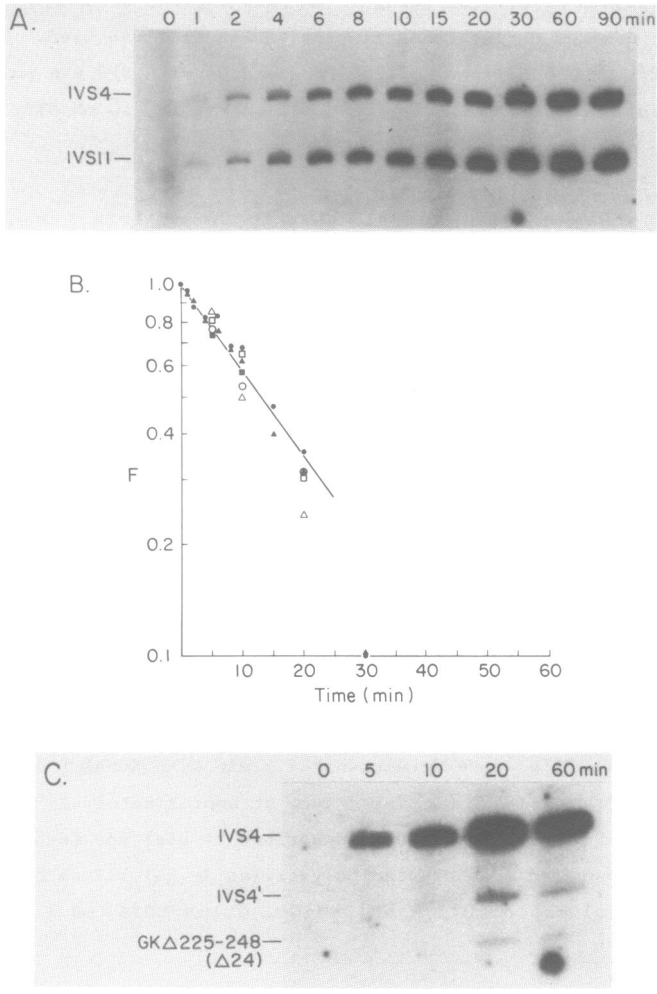

Figure 2. IVS Autoexcision Activities of Natural and Altered Forms of the T. thermophila Pre-rRNA Determined by the Amount of Covalent Addition of GTP to the 5' End of the IVS RNA.

(A) RNA was transcribed from an equimolar mixture of pIVS11 and pIVS4 templates under conditions inhibitory for self-splicing. The RNA was then incubated in splicing conditions in the presence of 2 μM α-^{32}P-GTP. At the indicated times after the addition of GTP, reactions were stopped. ^{32}P-labeled reaction products were subsequently analyzed by electrophoresis in a 4% polyacrylamide, 8 M urea gel and autoradiography.

(B) Rate of IVS excision in 2 μM GTP. F, the fraction of the IVS RNA that remains as part of the pre-rRNA, is calculated as $([IVS]_\infty-[IVS]_t)/[IVS]_\infty$, where $[IVS]_\infty$ is the maximum amount of IVS RNA that can be excised (determined from a 60 min reaction) and $[IVS]_t$ is the amount excised at time t (2). (o,●) pIVS4;

(Δ,▲) pIVS11; (□,■) pGKΔ233-246. Solid symbols, direct determination of ^{32}P-radioactivity by scintillation counting of gel slices. Open symbols, integration of peaks on spectrophotometric scan of the autoradiogram.

(C) Same as A except the RNA was transcribed from plasmid DNAs pGKΔ225-248, pIVS4 and pIVS4' present in molar ratios of 2:1:0.04. Plasmid pIVS4' arose from pIVS4 by deletion of 5 of the 6 Pst I linkers during propagation in E. coli; it therefore contains a 12 bp insert at the Bgl II site. A serial dilution of the transcription products confirmed that the ratio of excised IVS RNA from pIVS4 to that from pIVS4' was 1:0.04 (data not shown). Based on the autoradiogram shown here, the amount of IVS RNA excised from pGKΔ225-248 is approximately the same as that from pIVS4' even though the former was in 50-fold excess. We therefore calculate the efficiency of IVS excision of pGKΔ225-248 RNA in 2 μM GTP as ~2% of normal.

from pIVS4 was compared to the natural T. thermophila pre-rRNA transcribed from pIVS11 to determine the relative amount of IVS excision. The primary assay for IVS excision was to incubate uniformly labeled in vitro transcription products in splicing conditions. RNA was then separated on polyacrylamide gels to look for a species with the mobility predicted for an IVS with an insertion or deletion of known size. The pIVS4 RNA produced an IVS RNA ~50 nucleotides larger than that of pIVS11 (data not shown), indicating that the normal splice sites were being used. In a second assay pre-rRNA, synthesized by in vitro transcription with tritiated nucleotides, was incubated in splicing conditions in the presence of α-^{32}P-GTP, which becomes covalently added to the 5' end of the IVS RNA during splicing (1). Gel electrophoresis was then used to determine the size of the ^{32}P-labeled excised IVS RNA. Application of this procedure to an equimolar mixture of pIVS4 and pIVS11 RNAs is shown in Figure 2A,B. The same amount of IVS RNA was excised from the two precursors at each time point. Reactions were done at 2 μM GTP, well below the apparent K_m for this substrate (2), so that differences in either K_m or V_{max} would be detected. Therefore the 52 base insert in pIVS4 has no significant effect on the kinetic parameters of IVS excision.

Deletions at Position 236

A library of deletions was constructed at the same Bgl II site. Clones were characterized by restriction endonuclease analysis of the plasmids and dideoxy-sequencing of the RNA. Six cloned DNAs were transcribed and the RNA assayed for splicing activity. pIVS4 was included in each transcription as a positive control and as an internal standard to which the efficiency of IVS excision of the RNAs bearing the deletions could be compared. The IVS band was visible in the initial transcription products (Figure 3, odd numbered lanes). Further incubation in splicing conditions resulted in cyclization of a portion of the excised IVS and an increase in the total amount of excision (L IVS + C

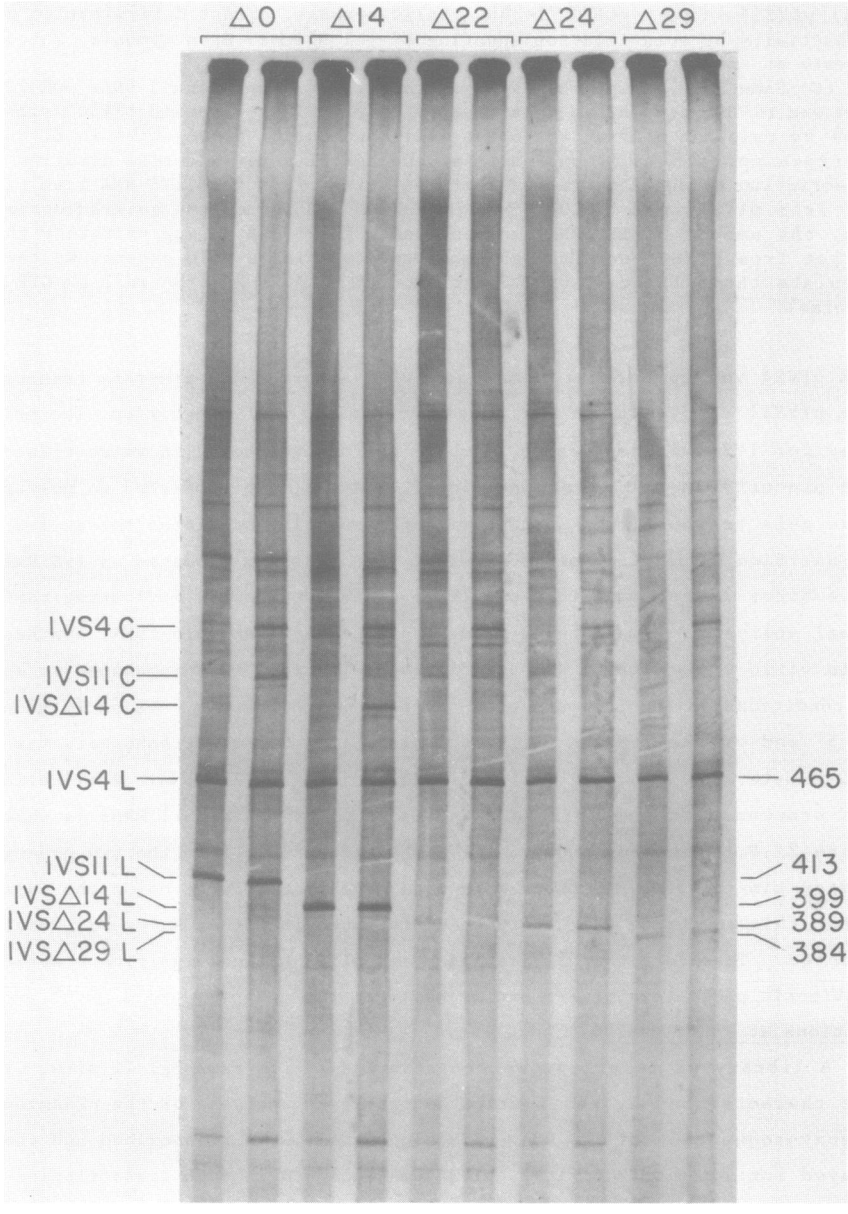

Figure 3. IVS Autoexcision Activities of Natural and Altered Forms of the Pre-rRNA Assayed Directly by Gel Electrophoresis. Each transcription reaction contained a mixture of two DNA templates. Plasmid pIVS4, containing a 52 bp insertion in the IVS of the T. thermophila rDNA, was included in each mixture as a positive control. The second template was either (Δ0) pIVS11, which has the natural sequence, or one of the deletion derivatives: (Δ14) pGKΔ233-246, (Δ22)

pGKΔ222-243, (Δ24) pGKΔ225-248, or (Δ29) pGKΔ224-252. pIVS11 and pIVS4 were mixed in a 1:1 ratio while each deleted plasmid was mixed in a 2:1 ratio with pIVS4. RNA was transcribed in the presence of ^{3}H-GTP, extracted with phenol/chloroform and precipitated with ethanol (left sample of each pair). A portion of each sample was incubated 30 min at 30° C in splicing buffer to increase the extent of IVS excision. Significant cyclization of the IVS also occurred during this additional incubation (right sample of each pair). Samples (20,000 cpm) were separated by electrophoresis in a 4% polyacrylamide, 8 M urea gel which was subjected to fluorography. The sizes indicated for the IVS RNAs are based on sequence information and not determined from the electrophoretic mobilities.

IVS) (even numbered lanes). These results were verified by the GTP end-labeling assay. Altered RNAs which retained splicing activity produced a discrete ^{32}P-labeled IVS. The results obtained with pGKΔ225-248 are shown in Figure 2C to give an indication of the sensitivity of this assay. The extent of IVS excision with pGKΔ225-248 RNA was less than 2% that of pIVS4.

The sequences and activities of deletions examined are summarized in Table 1 and Figure 4. The deletions fall into three broad categories. The first category includes constructs in which the deletion has little or no effect on splicing activity ("++" in Table 1); deletions including nucleotides 233-246 fall into this category. In the second category (deletions between nucleotides 224 and 252), splicing activity is significantly reduced but not eliminated ("+" in Table 1). In the third category (deletions to position 218 in a 5' direction or 258 in a 3' direction), splicing activity is undetectable ("-" in Table 1). In pGKΔ222-243 the inactivity could be due either to the deletion or to an A to C transversion at position 218.

Splicing of pGKΔ233-246 RNA was analyzed more quantitatively. As shown in Figure 2B, both the rate and the final extent of IVS excision were indistinguishable from those of the unaltered RNA from pIVS11. IVS excision from pGKΔ225-248 and pGKΔ224-252 RNAs was very inefficient at all GTP concentrations tested (2, 50 and 200 μM). It therefore appears that these IVS RNAs have either a reduced V_{max} of splicing or a reduced final extent of reaction, rather than an increased K_m for GTP.

Deletions at Position 42

Another library of deletions, centered at an endonuclease Sph I cleavage site at position 42 of the IVS, was constructed. Deletions were characterized and assayed for IVS excision as described above for the deletions at the Bgl II site. For these constructs an additional assay was used. The primer used for sequencing the deleted region allowed the sequence to be read through the 5' splice site. RNA was incubated in splicing conditions before sequencing. If excision occurred a strong stop was observed at the 5' end of the IVS in

Table 1. Sequences and Activities of Deletion Derivatives of the T. thermophila Pre-rRNA

	220 230 240 250 260		Activity
	. . Bgl II . . .		
pIVS11	GCCAAGUCCUAAGUCAACAGAUCUUCUGUUGAUAUGGAUGCAGUUC	Δ0	++
pGKΔ233-246*	GCCAAGUCCUAAGUCAAC.....GAUGGAUGCAGUUC	Δ14	++
pGKΔ222-243*	GCCCAGU.....UGAUAUGGAUGCAGUUC	Δ22	-
pGKΔ225-248	GCCAAGUCCU.....UGGAUGCAGUUC	Δ24	+
pGKΔ232-258	GCCAAGUCCUAAGUCAA.....UC	Δ27	-
pGKΔ218-244	GCC.....GAUAUGGAUGCAGUUC	Δ27	-
pGKΔ224-252	GCCAAGUCC.....UGCAGUUC	Δ29	+
	30 40 50 60 70		
	. Sph I . . .		
pIVS11	GGGAAAAGUUAUCAGGCAUGCACCUGGUAGCUAGUCUUUAAACCAA	Δ0	++
pESΔ41-44	GGGAAAAGUUAUCAGG....CACCUGGUAGCUAGUCUUUAAACCAA	Δ4	++
pESΔ40-44	GGGAAAAGUUAUCAG.....CACCUGGUAGCUAGUCUUUAAACCAA	Δ5	++
pESΔ41-46	GGGAAAAGUUAUCAGG.....CCUGGUAGCUAGUCUUUAAACCAA	Δ6	++
pESΔ35-44	GGGAAAAGUU.....CACCUGGUAGCUAGUCUUUAAACCAA	Δ10	+
pESΔ36-46	GGGAAAAGUUA.....CCUGGUAGCUAGUCUUUAAACCAA	Δ11	+
pESΔ39-49	GGGAAAAGUUAUCA.....GGUAGCUAGUCUUUAAACCAA	Δ11	+
pESΔ28-58	GGG.....UCUUUAAACCAA	Δ31	-

The top line in each section gives the natural sequence found in Tetrahymena thermophila. Each line below shows the sequences remaining in a deletion. The numbers in the plasmid name correspond to the numbers of the deleted nucleotides. Position 1 is the guanosine which is added to the 5' end of the IVS during splicing. (++) Approximately normal activity; (+) Substantially reduced activity; (-) Undetectable activity (<2% of normal). *Single base substitutions in addition to deletions.

all four dideoxynucleotide lanes as well as the control lane with no dideoxynucleotides (Figure 5A). If RNA from a deletion construct was unable to perform autoexcision, the sequence was easily read through the 5' splice site and into the exon (Figure 5B). The seven deletions at this site can also be grouped into three categories, as summarized in Table 1 and Figure 4. Deletion of nucleotides 40-46 has no apparent effect on the activity of the molecule. Deletions removing nucleotides 35-49 do not remove any essential structures but they do reduce the efficiency of splicing. One deletion which removed nucleotides 28-58 is inactive.

Insertions at Positions 135 and 409

Additional unique restriction sites were generated in pIVS11. Bam HI linkers (CGGATCCG) were inserted into pIVS11 which had been partially digested with restriction endonuclease Rsa I. Three insertion derivatives of the IVS were identified by restriction analysis. One of these (pJK39) contains three Bam HI linkers at position 135. The other two contain insertions at position 409. They differ in that pJK15 contains one Bam HI linker at this site, while

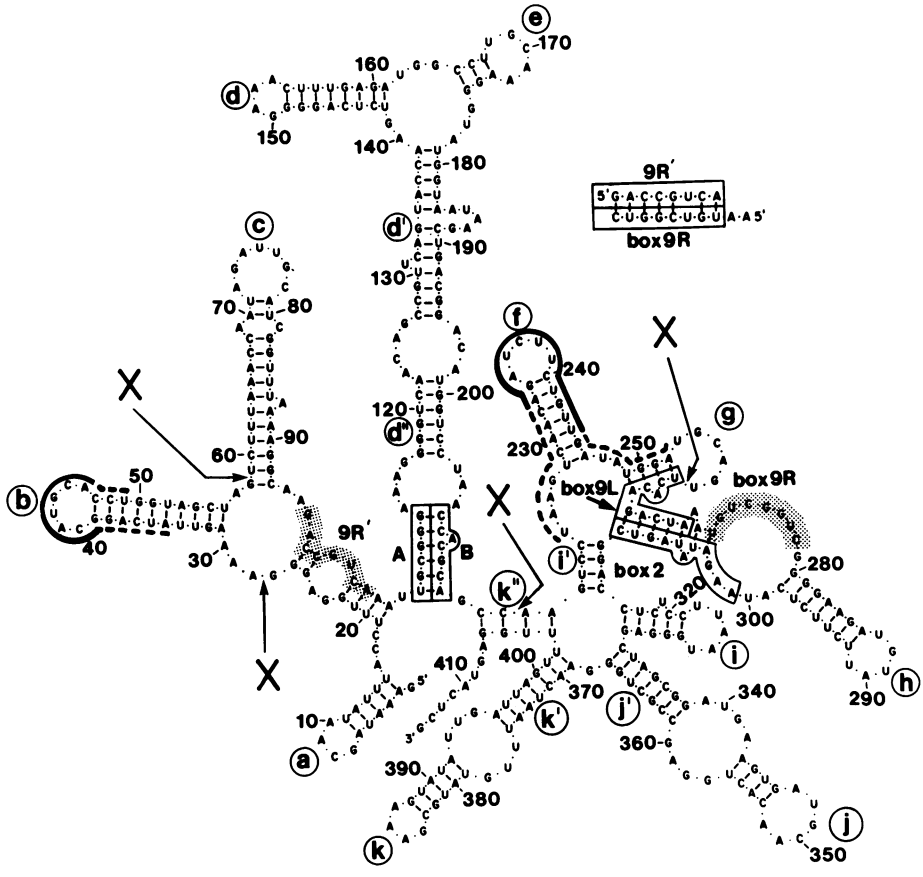

Figure 4. Structural Elements Required for IVS Activity. (———) Regions which can be deleted with little effect on excision activity. (- - - -) Regions whose deletion reduces but does not eliminate activity. (X) Endpoints of deletions for which no activity can be detected. The shaded and boxed regions have been hypothesized to be required for activity because they are conserved between fungal mitochondrial introns and the Tetrahymena rDNA IVS.

pJK29 contains 5 linkers. RNA from these plasmids was assayed by the end labeling method described above and in each case found to have a level of IVS excision activity comparable to RNA from pIVS11 (data not shown).

Because the inserts at position 409 are close to the 3' splice site, RNA from these plasmids was examined by several other methods. IVS RNA which was end-labeled with α^{32}P-GTP during splicing was gel-purified and incubated in cyclization conditions. The sample was then electrophoresed on a sequencing gel adjacent to a sample of RNA from pIVS11 which had been treated identi-

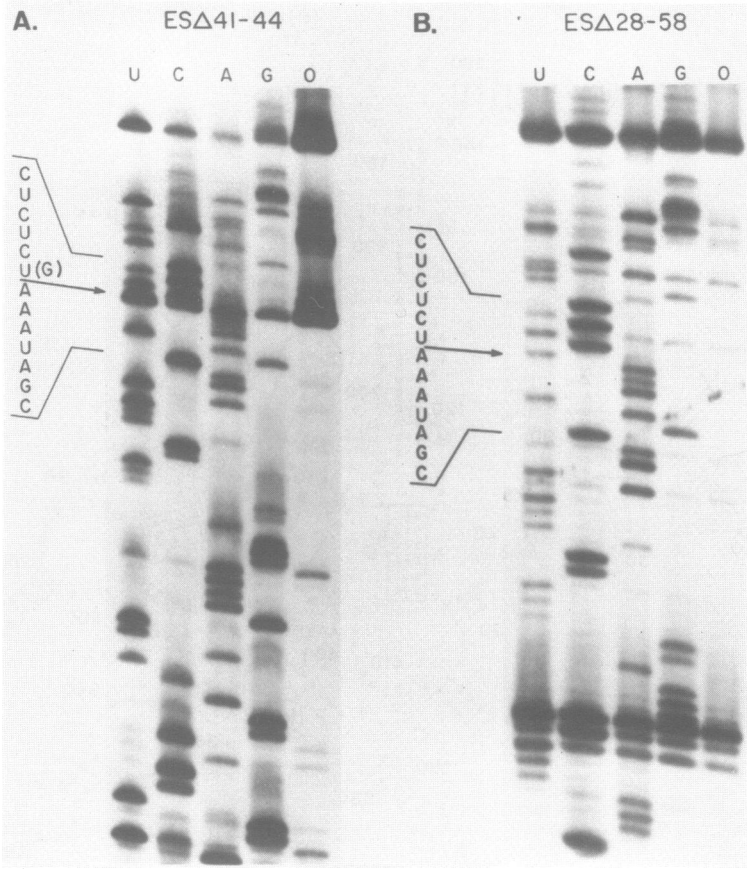

Figure 5. Primer Extension Assay for Autoexcision of Intervening Sequences with Deletions at Position 42. RNA was transcribed from (A) pESΔ41-44 and (B) pESΔ28-58. The RNA was incubated for 40 min at 30° C in splicing buffer. It was then hybridized to a ^{32}P-labeled synthetic oligonucleotide primer complementary to nucleotides 96-112 of the IVS. The primer was extended by reverse transcriptase (O) in the absence of any dideoxynucleotide, or (U,C,A,G) in the presence of a single ddNTP; the base complementary to the added ddNTP is indicated. Arrow indicates the 5' splice site. In the case of ESΔ41-44, the strong stop is one nucleotide beyond the splice site because of the G residue added to the IVS during splicing.

cally. Each IVS RNA was seen to release a ^{32}P-labeled oligomer which comigrated with the 15-mer released by pIVS11 RNA (data not shown). This provided the first indication that the normal 5' splice site and cyclization site were used. In addition uniformly labeled linear IVS RNA obtained from each insertion derivative was purified and tested for autocyclization. As shown in

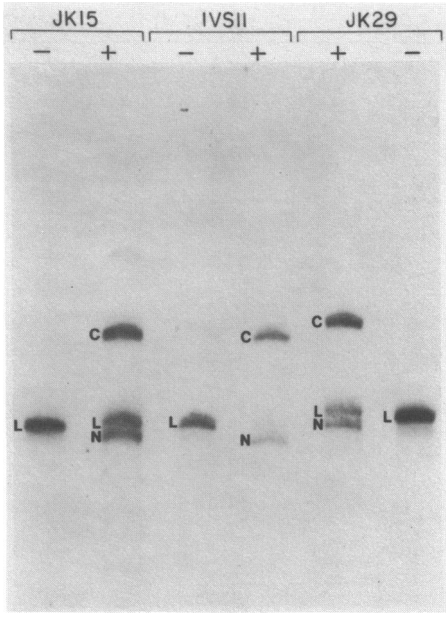

Figure 6. Cyclization of IVS RNA with Insertions at Position 409. Uniformly labeled linear IVS was gel-purified and (-) not incubated or (+) incubated 30 min at 42°C in splicing buffer. Samples were separated by electrophoresis in a 4% polyacrylamide, 8 M urea gel. The gel was dried and autoradiographed. (L) Linear IVS RNA; (C) Circular IVS RNA; (N) L-15 nicked circular IVS RNA. The natural IVS (IVS 11) was almost completely converted to the C and N forms after 30 min at 42°C, while conversion of the variant forms was incomplete; in a different set of RNA preparations, all three RNAs were completely cyclized (data not shown).

Figure 6, each linear IVS RNA was converted to the circular form and to the reopened circle (L-15 IVS, described in ref. 20). The linear, circular and L-15 forms of the IVS from pJK15 (1 linker inserted) were similar in size to those from the unaltered pIVS11; the corresponding IVS RNAs from pJK29 (5 linkers inserted) were somewhat larger.

In order to facilitate further analysis of the cyclization and splicing reactions, the Tetrahymena inserts of pJK15 and pJK29 were excised as 1.6 kb Hind III fragments and transferred to pSP62-PL (21) to create pJK15-SP6 and pJK29-SP6. A similar plasmid, pJK43-SP6, containing the natural IVS, was constructed by A. Zaug and used as a control. RNA was transcribed in vitro using SP6 RNA polymerase under conditions which inhibit splicing. The precursor RNA was gel-purified and incubated in the presence or absence of GTP. As shown in Figure 7, incubation in the presence of GTP led to the conversion

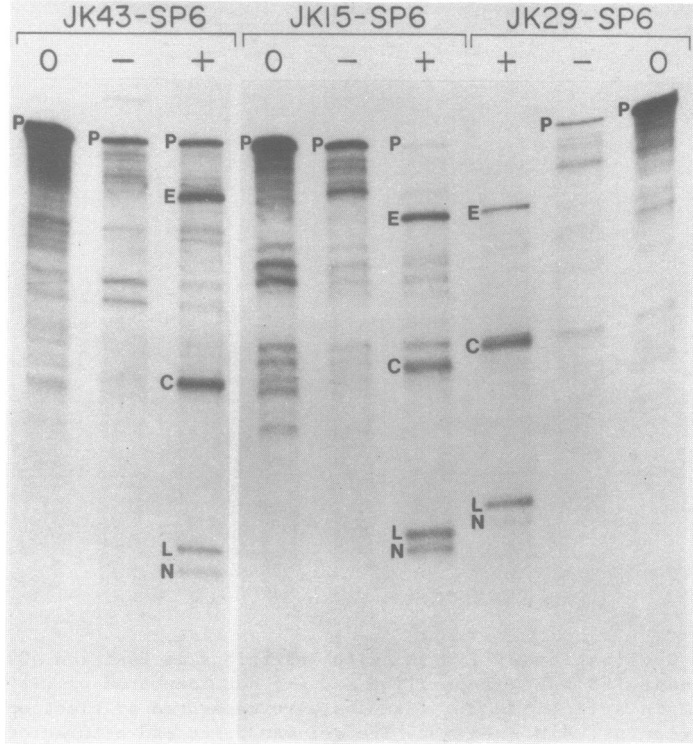

Figure 7. Splicing of Pre-rRNA with Insertions at Position 409. Plasmid DNAs were linearized with restriction endonucleases (pJK43-SP6 with Sma I and pJK15-SP6 and pJK29-SP6 with Eco RI) and transcribed in vitro with SP6 RNA polymerase. Uniformly labeled precursor RNA was gel-purified. (0) No further incubation; (-) Incubation at 30° C for 75 min in splicing buffer with GTP omitted; (+) Incubation under same conditions with GTP. Samples were separated by electrophoresis in a 4% polyacrylamide, 8 M urea gel. The gel was dried and fluorographed. (P) Precursor RNA; (E) Ligated exons; other RNA species indicated as in Figure 6. The different restriction sites used to linearize the plasmids give rise to ligated exons of different sizes, 1060 nucleotides for JK43-SP6 RNA and 927 nucleotides for JK15-SP6 and JK29-SP6 RNA.

of the precursor to RNA species corresponding in size to the ligated exons and to the three forms of the excised IVS RNA. Details of the identification of the band corresponding to the ligated exons will be presented elsewhere (T. Inoue, F. Sullivan, and T. Cech, manuscript in preparation).

The circular IVS RNA species were eluted from a preparative gel and sequenced by the dideoxynucleotide method using a synthetic primer complementary to nucleotides 67-83 of the IVS. One of the sequencing gels is shown in Figure 8. For all three RNAs, the sequence shows that the normal 3' end of

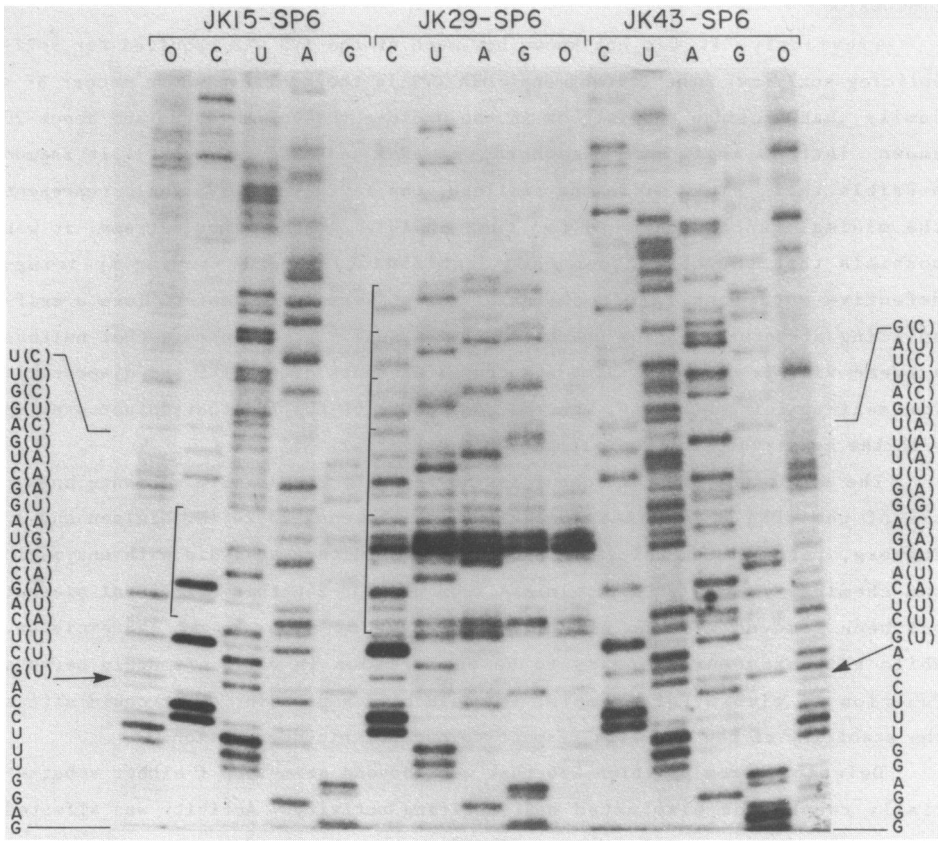

Figure 8. Sequence Analysis across the Cyclization Junction of C IVS RNA with Insertions at Position 409 Indicates that the Normal 3' Splice Site is Chosen. Gel-purified circular IVS was sequenced by extension of a ^{32}P-labeled synthetic primer, complementary to nucleotides 67-83 of the IVS, with reverse transcriptase in the presence of dideoxynucleotides. The nucleotide complementary to the ddNTP added to each sample is indicated. Brackets indicate the sequence of the Bam HI linkers. Arrows indicate the cyclization junctions.

the IVS is joined to the A at position 16. Thus, the insertion of extra nucleotides at position 409 changed neither the choice of the 3' splice site nor the choice of the cyclization site. In all three cases, the sequence of the circular IVS is superimposed on the sequence of the pre-rRNA, shown in parentheses in Figure 8. This is a result of the gel-purified circular IVS RNA being contaminated with products of pre-rRNA breakdown (note background in Figure 7). The intensity of the bands corresponding to the sequence of the precursor is greater than we had expected; it is possible that priming may be more efficient on the pre-rRNA than on the circular IVS RNA template.

DISCUSSION

Previously, it was not known how much of the IVS was required for self-splicing activity. The Tetrahymena rDNA IVS is the smallest known member of a family that includes two introns in the nuclear rDNA of Physarum and about 20 known introns in fungal mitochondrial rRNA and mRNA genes (22). It seemed possible that, since it is the smallest, the Tetrahymena IVS could represent the minimal functional size for such an IVS. At the other extreme, it was possible that the box sequences, identified by the cis-acting splicing-defective mutations in mitochondrial genes, were sufficient to form a self-splicing structure. Based on the deletion analysis, it appears that neither extreme view is correct. Some structural elements of the IVS are dispensible for self-splicing activity, whereas some parts of the IVS that do not overlap with the conserved sequence elements are essential.

The existence of stem-loop f is indicated by comparative sequence analysis of the rDNA from three species of Tetrahymena (23,24; H. Nielsen and J. Engberg, personal communication) and by RNA structure analysis with enzymatic and chemical probes (9,19). In RNA from pGKΔ233-246 this structural element has been removed with no effect on the rate or accuracy of IVS excision. While this structure appears to be superfluous in vitro, it could serve a function in vivo. For example, it could bind a protein, or it could affect the stability of the IVS with respect to nucleolytic degradation.

Deletions from position 236 that went beyond stem-loop f either substantially reduced or eliminated self-splicing activity. Activity was affected before the deletion end-points impinged on the conserved sequence elements B and box9L. From Table I it can be seen that the loss of activity did not correlate with the size of the deletion. For example, pGKΔ224-252, the largest deletion tested, was active while three of the smaller deletions were inactive. The inactive constructs had less symmetric deletions, and therefore extended further in one direction or the other from position 236. The important nucleotides appear to be somewhere between positions 218 and 224 on the left and between 253 and 258 on the right. We conclude that these nucleotides are involved in some structures required for activity. The structures could be different from the elements i', k" and g shown in the model in Figure 4; there is no real evidence for those particular base-pairing interactions (19). An alternative structure for this region, proposed by Michel and Dujon (7), has nucleotides 214-223 partially paired with 250-259 so as to extend the f stem. It could be this structure that is eliminated by the deletions that substantially reduce self-splicing activity.

A stem-loop structure equivalent to element **b** in the Tetrahymena IVS is present in many but not all group I introns (10,11). When present, it is in the same location with respect to the 5' splice site and contains approximately the same number of base pairs, 8-11. Neither the stem nor loop is conserved in primary sequence. We found that deletions within the **b** loop that are predicted to result in replacement of the normal CAUGCA loop with GCAC, CAC, or GGCC had no discernible effect on splicing activity. Larger deletions that are predicted to significantly destabilize the **b** stem substantially reduced activity, while one deletion that eliminated the stem and a few adjacent nucleotides eliminated activity. Thus, efficient self-splicing of the Tetrahymena IVS requires a double-helical stem in this position, with no apparent requirement for a particular sequence in the hairpin loop. Based on the phylogenetic evidence, there may very well be a similar structural requirement for splicing of some other group I introns. Ribosomal RNAs contain stem-loop structures that are similarly conserved in location but not in nucleotide sequence (25).

The fact that deletions within the IVS far from the splice sites can eliminate splicing activity is direct evidence for the importance of the IVS in the splicing reaction. This is in contrast to the situation observed for nuclear messenger RNA precursors. In yeast nuclear mRNAs only one internal site, the "TACTAAC box" (located near the 3' splice site), is required for splicing (26-28). In rabbit β-globin pre-mRNA no sequences other than those near the splice sites are observed to be indispensible (29). These results and others (reviewed in ref. 30,31) indicate that the structure of the intervening sequence far from the splice sites is unimportant for the splicing of most nuclear pre-mRNAs.

The results with the various insertion derivatives of the IVS indicate significant flexibility in the structural requirements for self-splicing activity. Insertions at three sites within the IVS did not influence IVS excision. RNAs with insertions at positions 15 and 328 have recently been tested and were also found to undergo normal IVS excision (G. Morin, J. Price, G. Gottlieb and T. Cech, unpublished data). Many of the insertions consisted of multiple repeats of a short palindromic sequence. Because of this self-complementarity, each insertion probably folds into a small helix minimizing its effect on the structure of the IVS. In the cases of positions 15 and 409, the insertion of a single 8 nucleotide linker was also tested and found not to interfere with either IVS excision or exon ligation (Figure 7 and M. Been and T. Cech, unpublished data). Presumably the inserted sequences are simply

looped out in the folded molecule and do not interfere with formation of the structure required for splicing.

The activity of RNA with insertions at position 409 has implications for the mechanism by which the 3' splice site is chosen. Position 409 is only 5 nucleotides upstream from the 3' splice site. If the 3' splice site were chosen by its distance from some structure within the IVS, the insertion of nucleotides at 409 should have resulted either in a new 3' splice site within the Bam linker or in the elimination of splicing. Instead, the normal splice site is used. This suggests that sequences distal to position 409 are involved in some structure that designates the 3' splice site. The critical sequences could be in the 3' exon, as in the internal guide sequence model of Davies et al. (8,11), or they could involve the last 5 nucleotides of the IVS.

The fact that the splicing activity is so tolerant of insertions may have implications for the evolution of this group of intervening sequences. The mitochondrial group I introns usually contain large open reading frames. If DNA containing an open reading frame were inserted into a pre-existing intron, the splicing activity of the RNA transcript of the gene might very likely be undamaged. In the Saccharomyces, Neurospora and Aspergillus mitochondrial rRNA introns, the conserved sequence elements flank the open reading frame (10,11) as if it had been acquired secondarily by some insertion event. Thus, insertions may be a major mechanism for the creation of diversity within this class of introns.

ACKNOWLEDGEMENTS

We are grateful to Carol Cech, Cammy DeLuca-Flaherty and Dennis Prosen for providing E. coli RNA polymerase, Marv Caruthers and Jeri Beltman for synthetic oligonucleotides, Art Zaug for pJK43-SP6 and Michael Been for comments on the manuscript. This work was supported by NIH grant GM28039. T.R.C. was supported by a Research Career Development award from the National Cancer Institute, Department of Health and Human Services.

+Present address: Amgen Development Inc., Boulder, CO 80301, USA

§Present address: Baylor College of Medicine, Houston, TX 77030, USA

*Present address: Brown University, Providence, RI 02912, USA

REFERENCES

1. Cech, T.R., Zaug, A.J. and Grabowski, P.J. (1981) Cell 27, 487-496.
2. Bass, B.L. and Cech, T.R. (1984) Nature 308, 820-826.

3. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Cell 31, 147-157.
4. Grabowski, P.J., Zaug, A.J. and Cech, T.R. (1981) Cell 23, 467-476.
5. Zaug, A.J., Grabowski, P.J. and Cech, T.R. (1983) Nature 301, 578-583.
6. Burke, J. and RajBhandary, U.L. (1982) Cell 31, 509-520.
7. Michel, F. and Dujon, B. (1983) EMBO J. 2, 33-38.
8. Waring, R.B., Scazzochio, C., Brown, T.A. and Davies, R.W. (1983) J. Mol. Biol. 167, 595-605.
9. Cech, T.R., Tanner, N.K., Tinoco, I., Weir, B.R., Zuker, M. and Perlman, P.S. (1983) Proc. Natl. Acad. Sci. USA 80, 3903-3907.
10. Michel, F., Jacquier, A., and Dujon, B. (1982) Biochimie 64, 867-881.
11. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C. (1982) Nature 300, 719-724.
12. Weiss-Brummer, B., Holl, J., Schweyen, R.J., Rödel,G. and Kaudewitz, F. (1983) Cell 33, 195-202.
13. Garriga, G. and Lambowitz, A.M. (1983) J. Biol. Chem. 258, 14745-14748.
14. Garriga, G. and Lambowitz, A.M. (1984) Cell 39, 631-641.
15. Tabak, H.F., Van der Horst, G., Osinga, K.A. and Arnberg, A.C. (1984) Cell, in press.
16. Schleif, R.F. and Wensink, P.C. (1981) Practical Methods in Molecular Biology, pp. 134-135, Springer-Verlag, New York.
17. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
18. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295-298.
19. Inoue, T. and Cech, T.R. (1985) Proc. Natl. Acad. Sci. USA 82, in press.
20. Zaug, A.J., Kent, J.R., and Cech, T.R. (1984) Science 224, 574-578.
21. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucleic Acids Res. 12, 7035-7056.
22. Waring, R.B. and Davies, R.W. (1984) Gene 28, 277-291.
23. Wild, M.A. and Sommer, R. (1980) Nature 283, 693-694.
24. Kan, N.C. and Gall, J.G. (1982) Nucleic Acids Res. 10, 2809-2822.
25. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-410.
26. Langford, C.J. and Gallwitz, D. (1983) Cell 33, 519-527.
27. Pikielny, C.W., Teem, J.L. and Rosbash, M.(1983) Cell 34, 395-403.
28. Langford, C.J., Klinz, F.J., Donath, C. and Gallwitz, D. (1984) Cell 36, 645-653.
29. Wieringa, B., Hofer, E. and Weissmann, C. (1984) Cell 37, 915-925.
30. Sharp, P. (1981) Cell 23, 643-646.
31. Greer, C.L. and Abelson, J. (1984) Trends Biochem. Sci., 139-141.