Volume 13 Number 23 1985 Nucleic Acids Research

Sites of circularization of the *Tetrahymena* rRNA IVS are determined by sequence and influenced by position and secondary structure

Michael D.Been and Thomas R.Cech

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

Received 23 September 1985; Accepted 4 November 1985

ABSTRACT

The sequence of the cloned Tetrahymena ribosomal RNA intervening sequence (IVS) was altered at the site to which circularization normally occurs. The alterations caused circularization to shift to other sites, usually a nearby position which followed three pyrimidines. While a tripyrimidine sequence was the major determinant of a circularization site, both location of a sequence and local secondary structure may influence the use of that sequence. For some constructs circularization appeared to occur at the position following the 5' G, the nucleotide added to the IVS during its excision. Portions of the internal guide sequence (IGS), proposed to interact with the 3'exon were deleted without preventing exon ligation. Thus if the IGS-3'exon interaction exists, it is not essential for splicing in vitro.

INTRODUCTION

The Tetrahymena thermophila ribosomal RNA (rRNA) gene contains an intervening sequence (IVS or intron) of 413 base pairs, and an early step in the processing of the rRNA precursor is the excision of the IVS (1,2,3). Splicing of the pre-rRNA and a subsequent reaction, circularization of the excised IVS, proceed in the absence of protein (4); the catalytic sites are provided by the structure of the RNA. The cleavage and ligation reactions occur by a transesterification (phosphoester transfer) mechanism requiring no energy-donating cofactor (5,6). In the proposed reaction sequence, there is a specific interaction between guanosine and the RNA which positions the guanosine for nucleophilic attack at the 5' splice site (7). The guanosine becomes covalently attached to the 5' end of the IVS via a normal 3'-5' phosphodiester bond, and a free 3' hydroxyl group is generated at the 3' end of the 5' exon. A second nucleophilic attack at the 3' splice site by the end of the 5' exon transfers the phosphodiester bond at the 3' splice site to the end of the 5' exon and results in exon ligation with IVS release. Circularization of the IVS occurs with attack at the phosphate between nucleotides 15 and 16 of the IVS by the guanosine at the 3' end of the IVS.

As in the initial reaction, a 3' hydroxyl group is generated on the oligonucleotide that is released. Circularization can also occur, with a lower efficiency, to a secondary site located between nucleotides 19 and 20 (8,9).

Information regarding the selection of circularization sites should be useful in understanding the process of splice site selection and the mechanism of the splicing reaction for two reasons. First, both splicing and circularization involve a transesterification mechanism which is mediated by the structure of the IVS (4,6). Circularization of the linear IVS therefore represents a minimal system for examining the RNA-catalyzed transesterification reaction. Second, similarities in the sequence preceding the 5' splice site (UCU) and the primary and secondary circularization sites (UUU and CCU) suggest that selection of all these sites could occur by a mechanism involving the same binding site. In addition, there is a distinct advantage to studying the circularization site rather than the 5' splice site. Alterations of the IVS RNA that affect the selection of a reaction site could either change the catalytic activity of the molecule (altered active site) or change a portion of the molecule that acts as a substrate. If IVS excision and exon ligation occur normally, it may be assumed that the alteration has not affected catalytic activity, and the changes seen in the circularization reaction reflect changes in the substrate portion.

In the work reported here, we investigated the sequence preference and other requirements for circularization by introducing insertions and deletions at the circularization site of a cloned Tetrahymena IVS sequence. RNA transcribed in vitro was tested for splicing activity and for the ability of the excised linear IVS to circularize. In all cases where the pre-rRNA was capable of splicing, the excised IVS RNA was capable to some extent of circularization. The positions to which circularization occurred were identified. It was found that the major sites of circularization resembled the sequence at the wild-type site, in that a tripyrimidine sequence immediately preceded the site of attack by the 3' terminal guanosine. Both position and structure also appeared to play a role in the site selection.

MATERIALS AND METHODS

Nucleotides and Enzymes

Unlabeled nucleoside triphosphates were purchased from P-L Biochemicals, labeled nucleoside triphosphates from New England Nuclear, restriction enzymes from New England Biolabs, S1 nuclease and T4 polynucleotide kinase from New England Nuclear, and RNAases T1 and U2 from CalBiochem. E.coli RNA

polymerase was provided by C. Cech.

Oligonucleotides

HindIII and BamHI linkers were purchased from New England Biolabs. The BamHI linker sequence was CGGATCCG. Synthetic primer IP67-17, which is 17 nucleotides long and complementary to the IVS sequence from position 67 to 84, was provided by J. Beltman and M. Caruthers.

Plasmid construction

Plasmids used in this study were derived from pTT1A3 (provided by J. Price). pTT1A3 has the IVS-containing 482 base pair ThaI fragment of Tetrahymena rDNA, isolated from pIVS11 (4), inserted into the HindIII site of pPlac (4). Transcription from the lac UV5 promoter in this plasmid generates wild type IVS RNA capable of in vitro splicing. To generate plasmids with insertions at the circularization site, pTT1A3 was partially digested with MnlI and unit-length linear molecules were isolated by preparative agarose gel electrophoresis. These were ligated in the presence of BamHI linkers, and plasmids from individual transformants of HB101 were screened for unique BamHI sites at the circularization site (Morin, Price, and Cech, unpublished). The DNA from one such plasmid (pGM342) appeared to initially have 6 BamHI linkers at this site, but due to spontaneous deletions, this preparation of DNA contained a mixed population of plasmids containing from one to six linker sequences at the same site. Upon transformation of this DNA, two stable clones containing one (pGB1) and three (pGB3) linker sequences were isolated. DNA sequence analysis of these two plasmids confirmed the presence of the linker sequences at the circularization site but indicated that a single nucleotide at the MnlI cut site had been deleted (see results). Deletions were generated from the BamHI site by cutting with BamHI and treating with nuclease S1 for 1, 5, 15, and 30 minutes. The plasmid DNA was then ligated, recut with BamHI and transformed into HB101. Miniprep plasmid DNA from individual transformants was screened for loss of the BamHI restriction site, and the size of the deletion was determined by sequencing the miniprep DNA.

Plasmid preparation

Plasmid-containing strains were grown in M9 medium plus casamino acids and the plasmid amplified with chloramphenicol. Plasmid was prepared and purified as described by Price et al. (10). Miniprep plasmid DNA was prepared by the boiling method of Holmes and Quigley (11) with some modifications (10).

DNA sequencing

The DNA sequence at the site of linker insertion in pGB1 and pGB3 was initially determined using the method of Maxam and Gilbert (12). The sequences around the circularization sites in all other plasmids were determined using primer extension. Approximately 0.5 µg of miniprep plasmid DNA was cleaved with HindIII, which cuts outside the region of interest, and the DNA annealed with 1 pmole of 5' end labeled primer (IP67-17) by heating to 95° for 5 minutes and placing on ice. Primer extension by reverse transcriptase in the presence of dideoxynucleotides, and sequencing gels were as described (13).

In vitro transcription

Uniformly labeled precursor RNA was transcribed from a gel-purified TaqI fragment of the plasmid DNA using E.coli RNA polymerase and [α-^{32}P]GTP (4). The reaction was continued for 60 minutes, terminated by addition of EDTA to 20 mM and extracted once with buffered phenol and once with chloroform. The nucleic acids were ethanol-precipitated and resuspended in 80% formamide containing 1mM EDTA and tracking dyes. The RNA was fractionated on a 4% polyacrylamide gel containing 50% urea (wt/vol). RNA of precursor size was recovered from a gel slice by soaking in 0.6 M ammonium acetate, 1 mM EDTA, and 1% SDS. RNA was ethanol precipitated twice, rinsed with cold 80% ethanol, dried, resuspended in 0.1 mM EDTA and frozen. RNA to be end-labeled in a splicing reaction was transcribed from supercoiled plasmid in a similar reaction that was scaled up 10 fold, and in which [^{3}H]GTP replaced the ^{32}P label. RNA was separated from unincorporated nucleotides by gel filtration over Sephadex G50-150, ethanol precipitated, resuspended in 0.1 mM EDTA and frozen.

RNA splicing conditions

In vitro splicing conditions were 30 mM Hepes (pH 7.5), 200 mM NaCl, 5 mM $MgCl_2$, and 0.1 mM GTP at 30° for 30 minutes. Splicing and circularization conditions were 30 mM HEPES (pH 7.5), 200 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM GTP at 42° for 30 minutes. Reactions were terminated by addition of EDTA to 20 mM.

To prepare IVS RNA end-labeled with ^{32}P, ^{3}H-transcripts were prepared and, after separation from unincorporated ribonucleoside triphosphates, the ^{3}H-RNA was spliced in the presence of 1 µM [α-^{32}P]GTP (600 Ci/mmole) to label the 5' end of the excised IVS (4). The IVS was then purified by denaturing polyacrylamide gel electrophoresis.

RNA sequencing and sizing oligonucleotides released on circularization Sequencing of end-labeled IVS RNA by partial nuclease and alkali cleavage was

as previously described (14). All samples were heated to 95° in 5 M urea prior to loading onto the gel. The products of circularization reactions were ethanol precipitated in the presence of carrier tRNA and resuspended in tracking dyes containing 5 M urea. The sequence ladders generated by ribonuclease and alkali consist of fragments with 3' phosphate groups, while those generated by circularization contain 3' hydroxyl groups and therefore migrate slower. The sequence of the 15-mer released from the wild type sequence had been determined previously (6) and the mobility shift due to charge differences in this size range could be accurately estimated. The difference in mobility is equivalent to approximately one nucleotide for the sizes of fragments being analyzed. To identify the size of the large oligonucleotide released by pGB3 IVS RNA, the oligonucleotide was isolated from a preparative gel, partially cleaved with alkali and rerun in a 20% polyacrylamide sequencing gel. The absence of an extra space between the full size oligonucleotide and the largest cleavage product indicated that it was running true to its size under these gel conditions despite the charge difference, and an accurate assignment of the size could be made.

RESULTS

Insertions at the circularization site

The plasmids used in these studies were derived from pTT1A3 by inserting BamHI linkers into an MnlI site which coincides with the major site of circularization (see Materials and Methods). Two resulting plasmids were sequenced. One (pBG1) had a single BamHI linker and the other (pGB3) had three linkers inserted at the location of the original circularization site. In both plasmids there was a deletion of the A which occurs at position 16 in the wild type sequence, the nucleotide to which circularization normally occurs (Figure 1).

The effects of the insertions were examined by analyzing the products of the precursor RNA spliced in vitro. From each plasmid to be tested, a Taq I restriction fragment containing the lac UV5 promoter and pre-rRNA sequences was transcribed in vitro with E.coli RNA polymerase in the presence of [α-^{32}P]GTP. The products were separated electrophoretically in a denaturing polyacrylamide gel, and RNA from the region containing full-length precursor was isolated. The RNA precursor had 5' and 3' exons (77 and 183 nucleotides, respectively) that contained both plasmid and Tetrahymena sequences. Splicing was done under two sets of conditions, one which minimizes circularization (5mM Mg^{2+}, 30°), and one in which both splicing and circularization are

```
pTT1A3
-10       -1 2          12          22          32
  ATGACTCTCT AAATAGCAAT ATTTACCTTT GGAGGGAAAA GTTATCAGGC

pGB1 [pGB3]
-10       -1 2          12                22          32
  ATGACTCTCT AAATAGCAAT ATTTCGGATCCGCCTTT GGAGGGAAAA GTTATCAGGC

pBS113
-10       -1 2          12             22          32
  ATGACTCTCT AAATAGCAAT ATTTCGCGCCTTT GGAGGGAAAA GTTATCAGGC

pBS153
-10       -1 2          12           22          32
  ATGACTCTCT AAATAGCAAT ATTTCGCCTTT GGAGGGAAAA GTTATCAGGC

pBS156
-10       -1 2                       22          32
  ATGACTCTCT AAATAG     (9)    CCTTT GGAGGGAAAA GTTATCAGGC

pBS157
-10       -1 2                       22          32
  ATGACTCTCT AAATAG     (11)     TTT GGAGGGAAAA GTTATCAGGC

pBS159
-10       -1 2                       22          32
  ATGACTCTCT AAATAG     (14)         GGAGGGAAAA GTTATCAGGC

pBS155
-10       -1 2                       22          32
  ATGACTCTCT AAATAGCAAT AT    (8)    GGAGGGAAAA GTTATCAGGC

pBS306
-10       -1 2          12                       32
  ATGACTCTCT AAATAGCAAT AT       (12)     GAAAA GTTATCAGGC
```

Figure 1. Insertions and deletions at the IVS circularization site.
The numbering of the IVS DNA sequence starts with 2 so that it will correspond with the numbering of the IVS RNA sequence (14) which contains the extra G at the 5' end. The underlined bases are linker sequences. In parentheses are the number of bases deleted from the wild-type sequence. In pGB3 there are three repeated linker sequences.

more efficient (10 mM Mg^{2+}, 42°).

Reactions of precursor containing the wild type IVS (pTT1A3) are shown in Figure 2, lanes 1-3. The major products are linear IVS (L IVS, 414 nucleotides), circular IVS (C IVS) as previously identified (6), and ligated exons (LE, 260 nucleotides). The 260 nucleotide band was identified as ligated exon sequences by two dimensional fingerprint analysis (Price and Cech, unpublished). Site-specific hydrolysis of the C IVS generated a linear molecule lacking the first 15 nucleotides of the 5' end (L - 15) (8). Possible origins of some of the other products are described in the legend to Figure 2.

When precursor RNA from each construct with an insert at the cir-

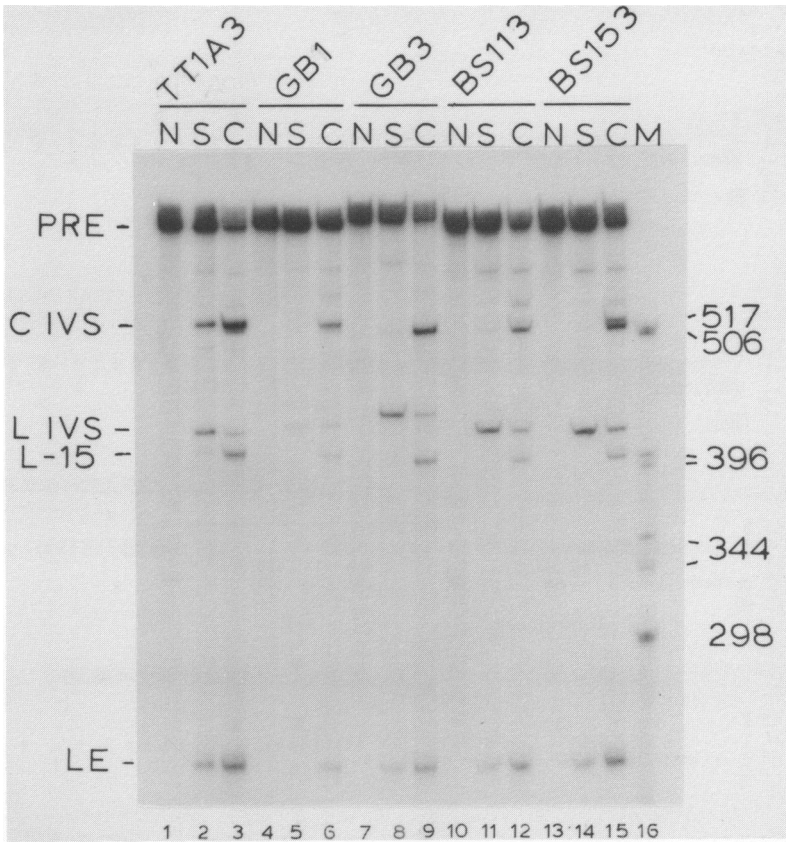
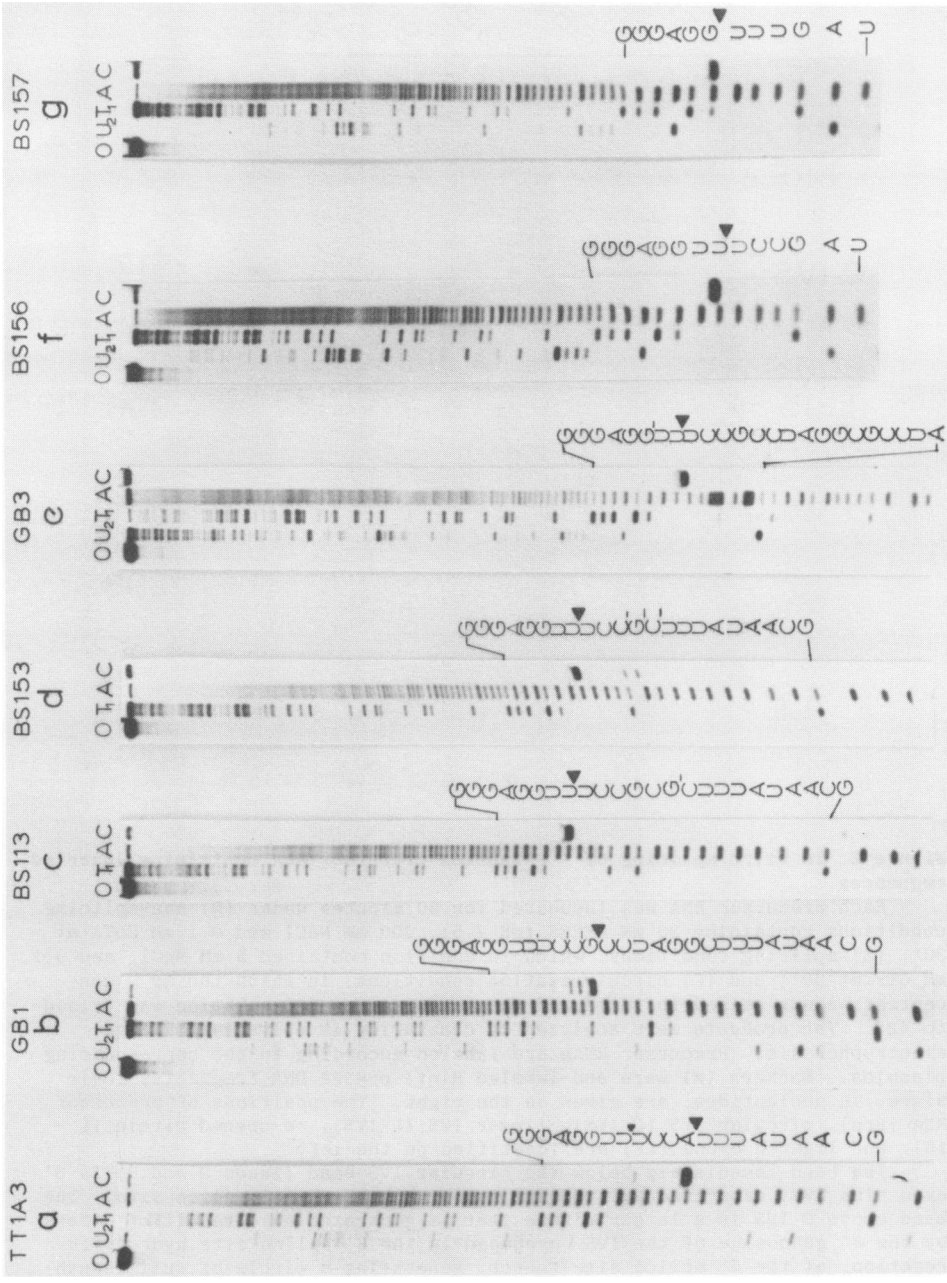


Figure 2. In vitro splicing of ^{32}P-labeled precursor RNA containing inserted sequences.

Each precursor RNA was incubated for 30 minutes under (N) non-splicing conditions containing 30 mM HEPES (pH 7.5), 200 mM NaCl and 0.1 mM EDTA at 30°; (S) splicing conditions, which in addition contained 5 mM $MgCl_2$ and 0.1 mM GTP at 30°; and (C) circularization conditions, in which the Mg^{+2} concentration was raised to 10 mM and the temperature of incubation was raised to 42°. The products were analyzed by denaturing 4% polyacrylamide gel electrophoresis. Precursor RNAs are labeled according to the corresponding plasmids. Markers (M) were end-labeled HinfI pBR322 DNA fragments; their sizes, in nucleotides, are given on the right. The positions of precursor RNA (Pre), circular IVS (C IVS), linear IVS (L IVS), re-opened circle (L - 15), and ligated exons (LE) are identified on the left.

The band immediately below the circular IVS band (lanes 2 and 3) is 5' exon plus IVS, generated by hydroxyl ion attack at the 3' splice site. The band above C IVS is a larger circle that is generated as a result of attack by the 3' guanosine of the IVS, produced in the 3' splice site hydrolysis reaction, at the 5' splice site thereby generating a circle of full length IVS (Inoue, Sullivan, and Cech, manuscript in preparation).


BS157
g
BS156
f
GB3
e
BS153
d
BS113
c
GB1
b
T T1A3
a

cularization site was put under splicing and circularization conditions, IVS was excised and exons were ligated. IVS RNA from pGB1 and pGB3 (Figure 2, lanes 5 and 8) migrated slower than the wild type IVS. In each case circular forms were generated (lanes 6 and 9) and the circular forms had mobilities similar to the wild type circle. Therefore, disrupting the UA sequence at the primary circularization site did not prevent circularization. In each case linear IVS eluted from the gel and put under circularization conditions generated a band which migrated with the mobility of the C IVS (data not shown). This result was interpreted to indicate that the L IVS was precursor to the circular form.

To determine the site of circularization, a second type of analysis was performed. Purified 5' end-labeled IVS was incubated under circularization conditions and the products analyzed in a sequencing gel. The analysis of wild type IVS sequence is shown in Figure 3, panel a. Circularization (lane C) produces fragments of 15 and 19 bases, corresponding to circularization at the primary and secondary sites. End-labeled IVS RNA from pGB1 generated a major oligonucleotide of a size that corresponded to circularization to a site within the linker sequence (Figure 3, panel b). Three other fragments were also generated, one of which corresponded to a reaction at the wild type secondary site of circularization. IVS RNA from pGB3, containing three linker sequences, gave a different result. The major site in this case occurred at the wild type secondary site, and a minor site occurred 2 bases 3' to this (panel e). The positions of the circularization sites, based on the sizes of the oligonucleotides released, are summarized in Figure 4.

Deletions at the circularization site

The major site of circularization in RNA from pGB1 occurred at the same position, or distance from a defined position within the IVS, as the primary site of circularization. Therefore, distance from some position within the IVS could be the determining factor in site selection. However, the sequence UCC immediately precedes this site, consistent with a pyrimidine requirement or preference as discussed above. To investigate positional and sequence

Figure 3. Circularization sites of IVS RNA containing inserted sequences.
End-labeled IVS RNA was circularized and the products analyzed in a 12% polyacrylamide sequencing gel (lanes C). Size markers from the same RNAs were generated with mild alkali treatment (lanes A) and partial digestion with ribonucleases U2 and T1 under denaturing conditions. RNA was also incubated in nuclease buffer conditions in the absence of nuclease (lanes 0). Major sites of circularization are marked on the sequences with large arrowheads. Secondary sites are indicated by the small arrowhead for the wild type sequence and with dashes for the altered RNAs.

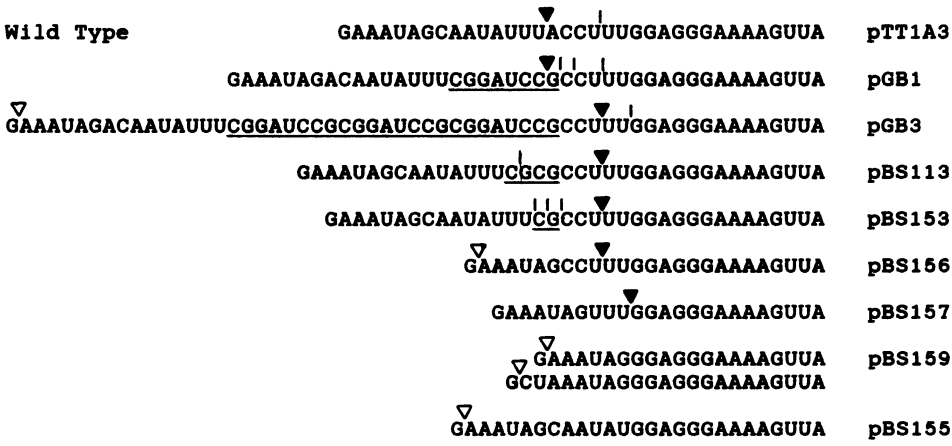


Figure 4. Locations of sites of circularization.
RNA sequences of the 5' region of the excised IVS and its derivatives are aligned with respect to an arbitrary internal point in the IVS. The sites of circularization are marked with arrowheads for the major site in each construct and with vertical lines for the secondary sites. Open arrowheads indicate release of the terminal GTP. It has not been determined whether GTP is released from both forms of the IVS derived from pBS159. Underlined nucleotides are inserted bases.

effects, pGB1 was cut at the unique BamHI site and treated with S1 nuclease to generate small deletions. The DNA sequences and the extent of the deletions for several of these constructs are shown in Figure 1.

RNA transcribed from these plasmids was tested for splicing activity. Removing part of the linker sequence (pBS113 and pBS153) did not inhibit splicing or circularization (Figure 2). Analysis of the major oligonucleotide that was generated under circularization conditions indicated that circularization occurred at the wild type secondary site rather than at the original position of the primary site (Figure 3). Minor sites were also apparent.

Four of the larger deletions (pBS156, pBS157, pBS159, and pBS155) also did not prevent splicing, and under circularization conditions a band that migrated at or near the position of circle was produced (Figure 5). When isolated linear IVS RNA was incubated under circularization conditions, it also generated a band running at the position of the circle (data not shown). This result was least expected for RNA from pBS159 and pBS155, because the entire pyrimidine-rich sequence around the circularization site had been deleted.

The sites of circularization were determined as before. For IVS RNA

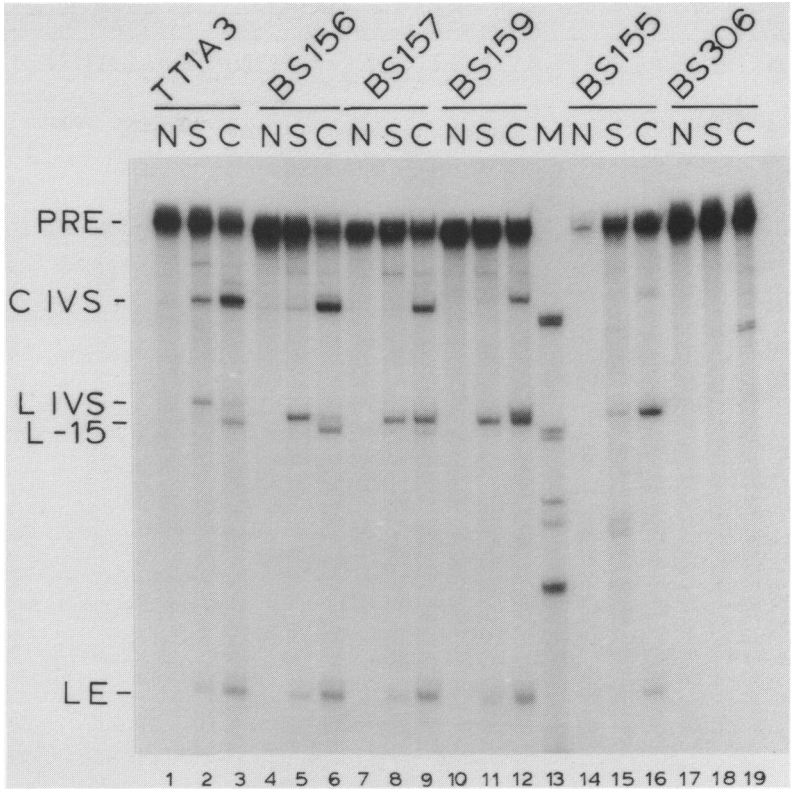


Figure 5. In vitro splicing of precursor RNA containing deletions around the circularization site.
Conditions are as described in the legend to Figure 2.

from pBS156, circularization again occurred at the secondary site. When the CC at this position was deleted (pBS157), circularization was shifted to a site following UUU (position 21 in the wild type sequence) (Figure 3). IVS RNAs from pBS159 and pBS155 gave two unexpected results. First, no oligonucleotides were produced under circularization conditions, even though it was evident from previous results that the excised IVS would circularize. However, a band co-migrating with GTP on the 20% polyacrylamide sequencing gel was generated (Figure 6). The product also chromatographed with GTP on PEI cellulose in two different solvent conditions (data not shown), and we conclude that it was GTP. Some of the other altered RNAs also produced a band co-migrating with GTP when analyzed on a 20% sequencing gel, but release of the 5' terminal GTP was not seen with the wild type IVS under these same

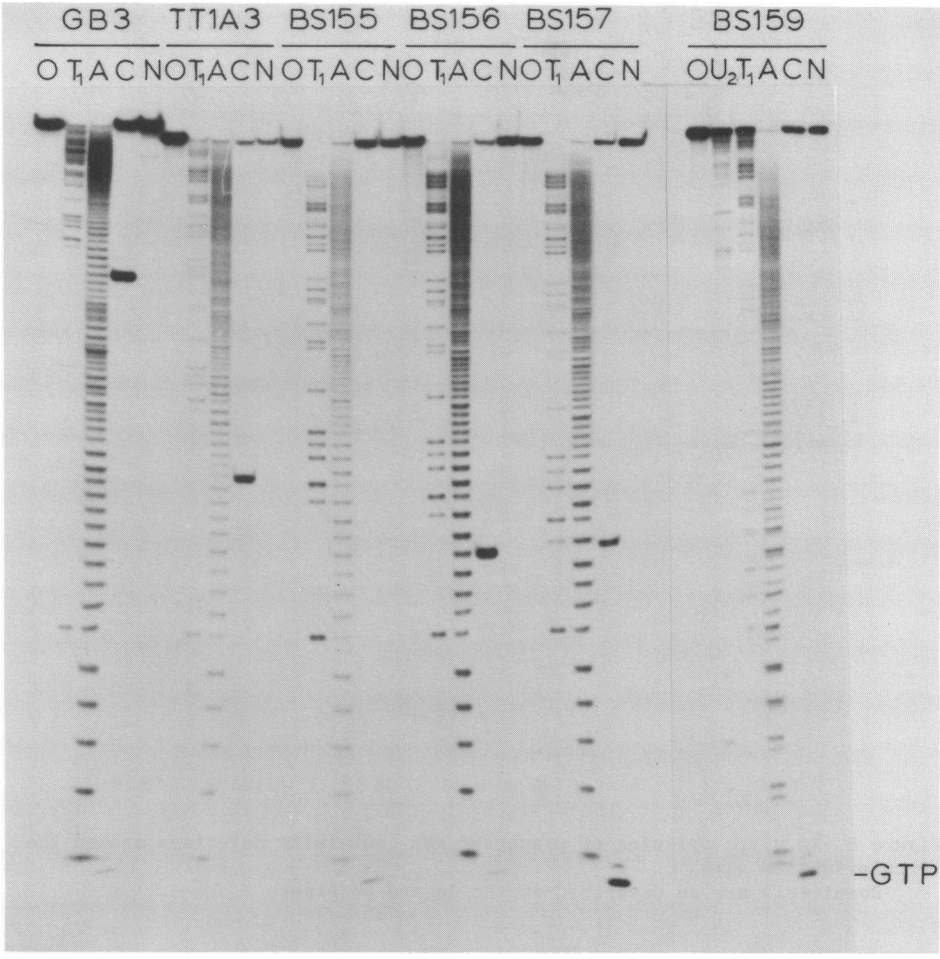


Figure 6. Circularization of end-labeled IVS and release of GTP.
Reaction conditions were as described in the legend to Figure 3. In the additional lane (N) the end-labeled IVS was incubated in circularization buffer minus Mg^{2+}. Products were analyzed on a 20% polyacrylamide sequencing gel.

conditions (Figure 6). Second, sequencing of the 5' end-labeled IVS RNA from pBS159 gave two overlapping sequences (Figure 6). Sequencing of the plasmid DNA gave an unambiguous and unique sequence in this region. When the end-labeled IVS was digested with RNAase A, two discrete labeled oligonucleotides were produced in similar amounts, one of which coincided with that produced by wild type (data not shown). It appeared that two different 5' splice

sites were being chosen during the splicing reaction by which the IVS was end-labeled. The two sites which are consistent with the sequence data are the wild type site (CUCU/A) and a site 2 nucleotides 5' to it (CUCU/CUA). Both sites follow a CUCU sequence. Whether one or both of these linear IVS RNAs released GTP under circularization conditions has not been determined.

RNA from the final construct (pBS306) did not splice (Figure 5, lanes 17, 18 and 19). The major product generated under circularization conditions ran as a doublet band at a position that corresponded in size to the 5' exon plus IVS. In the absence of GTP, only the lower of the two bands was generated (data not shown). Therefore, the lower band could have been generated by hydrolysis at or near the 3' splice site as described in the legend to Figure 2. The larger of the two bands, which migrated as though it were approximately 10 bases longer, is of unknown origin.

DISCUSSION

The IVS sequence was modified at the position where circularization occurs, and the effect upon the splicing and circularization reactions was studied. For most of these constructs, splicing was not prevented, and the site of circularization was shifted to one or more alternative positions. In a general way the sequence requirements near the reaction site resemble the situation found for nuclear mRNA splicing, where mutation of the normal site often reveal cryptic sites (15). Sequence, position, and secondary structure are obvious factors that could affect circularization site selection, and data in this paper provides evidence for effects by each.

Sequence requirement for circularization

For those cases where an oligonucleotide is released upon circularization, the major site follows 3 pyrimidines. In four of six cases the major site corresponded to the secondary site (CCU/U) in the wild type sequence (Table 1). In the two unique cases the sites were UCC/G and UUU/G. As one might expect, the minor sites had a less strict pyrimidine sequence requirement. There were two instances where a G was at position -1 and one where a G was at the -2 position. These results support the idea of a sequence preference for the reaction and suggest that a tripyrimidine binding site within the IVS is involved in specifying circularization sites.

Using a different approach, Sullivan and Cech (16) reached a similar conclusion. They studied the sequence dependence of a reverse circularization reaction in which short oligonucleotides (dimers and trimers) would cause the circle to reopen with addition of the oligonucleotide to the A16 of

Table 1. Sequences at circularization sites.

Major sites

positions	-4	-3	-2	-1	/	1	[a]
	A	U	U	U		A	[b]
	G	C	C	U		U	
	A	U	C	C		G	
	G	U	U	U		G	

Secondary sites

positions	-4	-3	-2	-1	/	1	
	A	C	C	U		U	[c]
	U	C	C	G		C	
	C	C	G	C		C	
	G	C	C	U		U	
	C	U	U	U		G	
	U	U	U	C		G	
	A	U	U	U		C	
	U	U	U	C		G	
	U	U	C	G		C	

[a] The phosphate being transferred is located between -1 and 1.
[b] The primary site in the wild type sequence.
[c] The secondary site in the wild type sequence.

the linear molecule. There was a strong preference for pyrimidine sequences in this reaction, with UCU showing the most reactivity. The linear IVS RNA, generated in the reopening reaction, could again circularize with release of the oligonucleotide that had been added. Thus, sequences 5' to the pyrimidine sequence were not required for the circularization reaction. This approach gave no information about positional effects or about sequence requirements 3' to the site, since that nucleotide is always an A.

Is there a sequence requirement 3' to the circularization site? In the wild type sequence, the tripyrimidine sequence at both the 5' splice site and the primary circularization site is followed by an A. Such was not the case for any other site (Table 1). Therefore, there is no absolute requirement at the +1 position. In the absence of additional data, such as site-specific base changes, the possibility that there may be some base preference has not been eliminated.

Position

A sequence of 3 pyrimidines was not the only requirement for selection of the circularization site. Position also appeared to play a role. While the distance of the circularization site from the 5' end of the IVS varied widely, its distance from positions internal to the IVS remained within a 9

nucleotide range. When the UUU sequence at position 13-15 in the wild-type sequence was shifted seven bases to the left in IVS RNA from pGB1, it was no longer used as a circularization site. Instead, circularization occurred primarily at a site within the linker sequence. However, when the same UUU sequence was located within one base of its original position (pBS153), it was used as a secondary site. The sequence UUU also occurred at position 19-21 in the wild type sequence and was used as a circularization site only in RNA from pGB3 and pBS157. For pBS157 RNA it was the only pyrimidine sequence available in the region.

Secondary structure

Multiple linker sequences (pGB3) caused circularization to occur at two sites outside of the linker sequences, rather at a site within the linker sequences as it did with IVS RNA from pGB1. An interpretation of this result is that the GC rich linker sequences are able to form a base-paired stem structure that prevents circularization within the linker sequence. The sequencing data (Figure 3) are consistent with a strong stem-loop structure; there was less cutting by ribonucleases T1 and U2 in the stem portion than in the loop portion of such a structure. In addition two areas of strong compression were observed in the alkaline ladder of the same gel. If base pairing prevents circularization to the site within the linker sequences, it would imply that circularization requires the site to be free to interact with another structure within the IVS. It is possible that some of the positional effects discussed above could be due to unidentified secondary structures that influence site selection in a similar manner.

The structure of the IVS proposed by Cech et al. (14) has the 5' terminal 15 bases forming a short stem-loop structure in which the three pyrimidines at the circularization site are base paired. The results from the insertion and deletion studies presented here show that the structure is not required for the circularization reaction and suggest that such a structure would inhibit circularization to that site.

GTP release

Circularization with release of GTP had not been seen previously. In two of the constructs (pBS159 and pBS155) circularization occurred and only the terminal GTP was released under circularization conditions. In another (pBS157) a significant amount of GTP was released along with an oligonucleotide. GTP release was not detectable with wild-type IVS under these conditions. It has not been shown that GTP release is dependent upon circularization. However, formation of a covalently closed circle in the

absence of an energy-donating cofactor requires that the total number of phosphodiester bonds be conserved, and therefore some bond must be broken. As a consequence, circularization of IVS RNA from pBS159 and pBS155 is presumably dependent on GTP release. As with the release of oligonucleotides in the other constructs, the released guanosine mononucleotide contained a 3' hydroxyl rather than a 3' phosphate; base catalyzed hydrolysis would have generated pppGp rather than pppG-OH. Circularization to a site following the 5' terminal G would appear to be fundamentally different from the other circularization reactions in which the transesterification involves a site within the RNA that is preceded by a tripyrimidine sequence. It is possible that this type of circularization is a guanosine exchange reaction in which the 5'-terminal G is exchanged for the 3'-terminal G, thereby forming a circle. Such an exchange reaction might involve activation by the IVS guanosine binding site (7).

Splicing, the internal guide sequence and circularization

All but one of the constructs studied were capable of splicing and IVS excision. In pGB3 the insertion of 23 nucleotides near the 5' end did not prevent what appeared to be normal in vitro activity. The deletion of up to 11 nucleotides between the 5' splice site and the interior of the IVS similarly did not affect choice of the 5' splice site, and only in the case of a construct with a larger (14 nucleotide) deletion was the 5' splice site not chosen accurately. Thus the 5' splice site is not chosen by its distance from a fixed point within the IVS, and is presumably chosen primarily by its nucleotide sequence. The 3' splice site also appeared to be chosen correctly in each case, as judged by the size of the ligated exons, but the exact position was not determined.

A sequence that can pair with the exons at the 3' and 5' splice site is frequently found in group I introns (17), of which the Tetrahymena pre-rRNA IVS is a member. Davies and his colleagues have proposed that this internal guide sequence (IGS) correctly aligns the 3' and 5' splice sites for the splicing reaction. The specific IGS structure they have proposed for the Tetrahymena IVS (18) is shown in Figure 7 (structure a). All of the constructs used in this study disrupt the portion of the IGS that would pair with the 3' exon. For example, RNA from pBS155 and pBS159 are still capable of exon ligation even though the 5' portion of the IGS has been deleted and replaced with quite dissimilar sequences. When structures similar to the IGS are drawn, the extent of pairing to the 3' exon is much less (Figure 7,structures b and c). Shifting the alignment of the 3' splice site a few

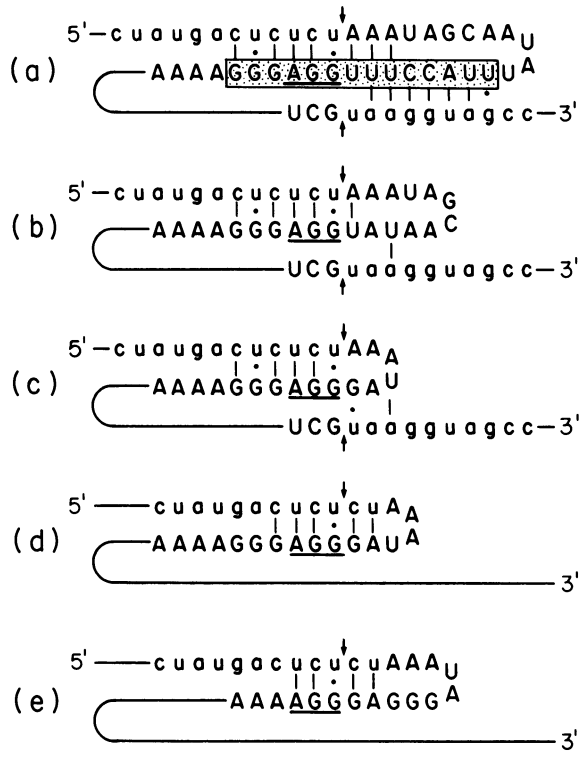


Figure 7. Hypothetical secondary structures which may be involved in splice site selection and activation.

Structure a is the IGS (internal guide sequence) according to Waring et al. (1983) for the wild type sequence. Structures b and c show RNA from pBS155 and pBS159 drawn in a similar manner. Structures d and e show alternative pairing possibilities of the 5' exon for RNA from pBS159. Exon sequences are shown in lower case letters, IVS in capital letters. Arrows indicate splice sites or potential splice sites. The underlined GGA sequence in each structure represents the proposed binding site for the tripyrimidine sequences that precede the 5' splice site and the circularization sites (see text).

bases either way did not greatly improve the strength of the interaction. An alignment mechanism, by which the exons are brought into proximity, would appear to be necessary for any splicing reaction. However, for the proposed IGS to play this role by itself would require that a variety of structures with nearly random base-pairing possibilities be allowed for 3' exon alignment. Thus, models in which the 3' splice site is positioned via other interactions, such as pairing the last few bases of the IVS, should be considered.

The possibility that the purine-rich region of the proposed IGS pairs with the exon at the 5' splice site is still attractive. This interaction has also been proposed by Michel and Dujon (19) for group I introns. In the proposed structures the UCU at the 5' splice site is paired with the sequence GGA (Figure 7, structure a). GGA was suggested by Sullivan and Cech (16) as a likely binding site for di- and tri-nucleotides which act to reopen the circular form of the IVS. In the studies presented here, RNA from a plasmid with a deletion that extends into this region (pBS306) appeared not to react at the 5' splice site, while other deletions of similar sizes were capable of splicing. Binding of the 5' exon to the purine-rich region can also account for the inaccuracy of 5' splice site selection in RNA from pBS159. As shown in structures d and e (Figure 7), alternative pairing possibilities for this RNA can be drawn. One of these pairings (stucture d) brings the UCU at positions -3 to -5 into the GGA site which is proposed to be normally occupied by the UCU at positions -1 to -3, (underlined GGA in structures 7a, b, and c), while the other pairing (structure e) brings the UCU into a new GGA site. The calculated stability for structure 7d is close to that of structure 7c, while that of structure 7e is slightly less. On the other hand, if similar alignments are made with the wild type sequence (not shown), there is a considerable loss in stability relative to that of the structure shown in 7a. Therefore, this mechanism for splice site selection would predict a higher probability of choosing an incorrect splice site with IVS RNA from pBS159 than with the wild type IVS. Whereas there was little evidence to support 3' splice site binding to the proposed IGS, these data can be interpreted as support of a mechanism of 5' splice-site selection based on base-pairing interactions with the IGS.

We have hypothesized that the 5' splice site and the circularization site are bound successively to a common active site in the IVS (16; Inoue, Sullivan and Cech, manuscript submitted). It is possible to draw a number of secondary structures in which the tripyrimidine sequence at the various circularization sites is paired to one or the other of the GGA sequences underlined in Figure 7. We propose that such an interaction is necessary for both guanosine attack at the 5' splice site and circularization.

Selection of both the 5' splice site of the Tetrahymena pre-rRNA and the circularization site of the excised IVS appear fundamentally similar to selection of 5' splice sites in nuclear pre-mRNAs, in that the sites are specified largely by the adjacent nucleotide sequences. In the rRNA case, the critical oligopyrimidine sequence is upstream from the reaction site,

while in pre-mRNA the critical GU$^{A}_{G}$AG is downstream from the 5' splice site. In rRNA, the 5' splice site and circularization site sequences probably bind to a sequence within the IVS. In pre-mRNA, 5' splice site binding to internal sequences has also been proposed (20,21), but it is likely that the most important interaction involves binding of the 5' splice site to U1 snRNA (22,23,24). In both the rRNA and mRNA systems, the RNA-RNA interactions required for splice site recognition appear to involve Watson-Crick basepairing.

ACKNOWLEDGMENTS

We thank J. Price and G. Morin for providing the two initial plasmids and for their insight and suggestions on this project. This work was supported by grants from the NIH (GM28039) and the American Cancer Society (NP-374). M.D.B. was supported by a Damon Runyon-Walter Winchell Postdoctoral Fellowship, and T.R.C. was supported by a Research Career Development Award from the National Cancer Institute.

REFERENCES

1. Wild, M.A. and Gall, J.G. (1979) Cell 16, 565-573.
2. Din, N., Engberg, J. Kaffenberger, W. and Eckert, W. (1979) Cell 18, 525-532.
3. Cech, T.R. and Rio, D.C. (1979) Proc. Natl. Acad. Sci. USA 76, 5051-5055.
4. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Cell 31, 147-157.
5. Cech, T.R., Zaug, A.J. and Grabowski P.J. (1981) Cell 27, 487-496.
6. Zaug, A.J., Grabowski, P.J. and Cech, T.R. (1983) Nature 301, 578-583.
7. Bass, B.L. and Cech, T.R. (1984) Nature 308, 820-826.
8. Zaug, A.J., Kent, J.R. and Cech, T.R. (1984) Science 224, 574-578.
9. Zaug, A.J., Kent, J.R. and Cech T.R. (1985) Biochemistry, in press.
10. Price, J.V., Kieft, G.L., Kent, J.R., Sievers, E.L. and Cech, T.R. (1985) Nucleic Acids Res. 13, 1871-1889.
11. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
12. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology. 85, 499-560.
13. Inoue, T. and Cech, T.R, (1985) Proc. Natl. Acad. Sci, USA. 82, 648-652.
14. Cech, T.R., Tanner, N.K., Tinoco, I.,Jr., Weir, B.R., Zuker, M. and Perlman, P.S. (1983) Proc. Natl. Acad. Sci USA 80, 3903-3907.
15. Mount, S. and Steitz, J. (1983) Nature 303, 380-381.
16. Sullivan, F.X. and Cech, T.R. (1985) Cell, 42, 639-648.
17. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A., and Scazzocchio, C. (1982) Nature 300, 719-724.
18. Waring, R.B., Scazzocchio, C., Brown, T.A. and Davies, R.W. (1983) J. Mol. Biol. 167, 595-605.
19. Michel, F. and Dujon, B. (1983) EMBO J. 2, 33-38.
20. Pikielny, C.W., Teem J.L., and Rosbash, M. (1983) Cell 34, 395-403.
21. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) Nature 313, 552-557.

22. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, M., and Steitz, J. (1983) Cell 33, 509-518.
23. Padgett, R.A., Mount, S.M., Steitz, J.A. and Sharp, P.A. (1983) Cell 35, 101-107.
24. Kramer, A., Keller, W., Appel, B., and Luhrmann, R. (1984) Cell 38, 299-307.