
The intervening sequence excised from the ribosomal RNA precursor of *Tetrahymena* contains a 5'-terminal guanosine residue not encoded by the DNA

Arthur J. Zaug and Thomas R. Cech

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

Received 3 March 1982; Revised and Accepted 12 April 1982

ABSTRACT

The ribosomal RNA precursor of *Tetrahymena thermophila* contains a 0.4 kilobase intervening sequence that is excised as a linear RNA molecule ("IVS RNA") and subsequently cyclized. *In vitro* transcription in isolated nuclei was used to accumulate the IVS RNA. IVS RNA labeled at its 5' end was subjected to sequencing gel analysis and terminal nucleotide analysis. In addition, uniformly labeled IVS RNA was cleaved with RNAase T₁, and the resulting oligonucleotides were studied by two-dimensional fingerprinting. The IVS RNA was found to be a unique molecule with no discernible terminal heterogeneity. The 5'-terminal nucleotide is a guanosine that is not present at the corresponding point in the DNA sequence, determined by N. Kan and J. Gall (see adjoining paper). This nucleotide is added to the IVS during splicing [Cech, Zaug, and Grabowski (1981) *Cell* 27, 487-496]. Based on the sequences near the ends of the RNA, the remainder of the RNA sequence is colinear with that of the DNA. The IVS RNA has 5'-monophosphate and 3'-hydroxyl termini. Comparison of these results to those obtained previously for yeast tRNA intervening sequences leads us to conclude that the splicing mechanisms are fundamentally different for these two classes of transcripts.

INTRODUCTION

Intervening sequences have been identified in many eucaryotic genes that code for mRNA, rRNA, and tRNA. In many cases, the splice junctions have been localized by comparing the sequence of the DNA to that of the RNA or the polypeptide encoded by the gene (reviewed in references 1-3). Additional information can be obtained by analyzing the primary structure of the intervening sequence product of splicing. If an intervening sequence is a unique molecule whose sequence is colinear with that of the DNA, it is likely that it is excised as a linear molecule and not as a circle that is subsequently broken. In addition, the location of phosphoryl and hydroxyl groups on the RNA termini gives indirect evidence about the specificity of the splicing enzymes. For example, the 3'-phosphoryl and 5'-hydroxyl ends of the yeast tRNA^{Tyr} intervening sequence led Knapp *et al.* (4) to propose a nuclease that creates 3'-phosphoryl ends and a ligase that joins a 3'-phosphate to a 5'-hydroxyl group.

The excised intervening sequences of yeast tRNA precursors are the only ones that have heretofore been analyzed at the nucleotide sequence level (4). We now provide sequence information about a second type of IVS, that excised from the ribosomal RNA precursor of Tetrahymena. This excised IVS RNA has a very short lifetime in vivo (~ 6 sec; S. Brehm and T. Cech, manuscript submitted). The RNA accumulates, however, when produced by in vitro transcription and splicing in isolated Tetrahymena nuclei (5) or nucleoli (6). The IVS initially produced by the splicing reaction is linear, but it is subsequently converted to a circular molecule (7). Isolated linear IVS RNA, purified by SDS-phenol extraction, denaturing gel electrophoresis, and extensive proteinase K treatment, has the intrinsic ability to cyclize itself when incubated at 37°C in Mg⁺²-containing solution (3). This autocyclization activity appears to be a property of the RNA molecule itself, although we have not been able to exclude the possibility that there is an unusually stable enzyme bound to the RNA.

In the work described here, we have focused on the structure of the linear form of the IVS RNA. We find that the RNA has 5'-phosphoryl and 3'-hydroxyl termini, opposite to the case of yeast pre-tRNA intervening sequences. We also find that the RNA has a 5'-terminal guanosine residue not encoded by the DNA. Addition of the guanosine occurs during the excision of the IVS from the pre-rRNA, and appears to be an essential step in the splicing mechanism (8).

MATERIALS AND METHODS

Growth of cells and isolation of nuclei

Tetrahymena thermophila strain B VII were grown in 2.8 l Fernbach flasks to densities of $1-2 \times 10^5$ cells/ml and nuclei were isolated as previously described (5).

Isolation of IVS RNA

Transcription in isolated nuclei was carried out as previously described (8) using 120 mM (NH₄)₂SO₄ to optimize the excision of the IVS and a temperature of 30°C to select for the linear form. Each 0.45 ml transcription mixture contained either 125 µCi [³H]UTP or 125 µCi each of the four [α-³²P] nucleoside triphosphates. After deproteinization and incubation with DNAase I, the IVS RNA was partially purified by centrifugation in a 15-30% sucrose gradient as described by Grabowski et al. (7), except that no 50% sucrose cushion was used. Circular IVS RNA was produced by incubating the purified IVS RNA for 20 min at 39°C in 0.01 M Tris-HCl (pH 7.5), 25 mM MgCl₂, 50 mM (NH₄)₂SO₄ ("autocyclization conditions"). Both linear and circular IVS RNA were puri-

fied to homogeneity by electrophoresis at room temperature in a 4% polyacrylamide, 8 M urea gel (8), the RNA being recovered from the gel at 4°C using the crush and soak procedure (9). From 2 l of Tetrahymena, 0.5 µg (3.7 pmoles) of pure IVS RNA was routinely isolated.

RNA sequence determination

End labeling of gel-purified linear IVS RNA followed the procedure of Donis-Keller *et al.* (9) with slight modifications. ³H-labeled linear IVS RNA (1.0 pmoles) was first treated with 40-50 U of bacterial alkaline phosphatase (BRL) at 37°C for 30 min in 100 mM Tris-HCl (pH 8.0), extracted 3x with phenol, precipitated with 70% ethanol, and dissolved in 10 µl of H₂O. End labeling was done at 37°C in 25 µl using 4 U of polynucleotide kinase (NEN) and 1 µM [γ -³²P]ATP (3000-7000 Ci/mmole). After ethanol precipitation, the labeled RNA was purified by gel electrophoresis as described above. End-labeled IVS RNA was sequenced according to published procedures (9,10) with slight modifications (8). Enzyme to substrate ratios (units per microgram of carrier tRNA) were 2×10^{-3} for T₁, 4×10^{-4} and 0.8×10^{-4} for U₂, and 4×10^{-2} for Phy M.

End analysis

Purified end-labeled linear IVS RNA was totally digested with RNAase T₂ for 1 hr at 37°C in 5-10 µl of 10 mM Tris-HCl, 1 mM EDTA (pH 7.5) using 5 U of enzyme. Nuclease P₁ digestions were done at 60°C using 2 U of enzyme. Products were analyzed by PEI-cellulose ascending chromatography and by high voltage paper electrophoresis (see legend to Figure 2). Nucleoside 3',5'-bisphosphates (pNp) markers were synthesized from the corresponding 3'-monophosphates (Sigma) using polynucleotide kinase and [γ -³²P]ATP. Nucleoside 5'-monophosphate markers were made by removing the 3'-phosphates of the pNp markers using nuclease P₁ as described above.

RNA fingerprinting

Uniformly labeled linear or circular IVS RNA was digested to completion with RNAase T₁ (Sankyo) at 37°C for 30 min using 1 µg of T₁ per 20 µg of carrier tRNA. Reaction mixtures were spotted 10 cm from the end of a 3 x 55 cm strip of cellulose acetate (Schleicher and Schuell). Electrophoresis was carried out at 5000 V in 5% acetic acid, 0.25% pyridine, 7 M urea (pH 3.5) for 50 min. The partially separated T₁ products were blotted onto a 20 x 40 cm glass-backed PEI plate, which was then washed 2x in 95% ethanol and chromatographed in pyridine formate (pH 3.5) using a modified version of the procedure of Griffin (12). After the first 10 cm of chromatography in 1.0 M pyridine formate, the plate was transferred to 2.2 M pyridine formate. Chromatography continued until the solvent front was 5-6 cm from the top of the plate. The

separated oligonucleotides were located by autoradiography. Each spot was scraped from the glass plate and washed with 95% ethanol to remove the urea. RNA was recovered by elution of the PEI with 2 M triethylamine bicarbonate.

Base composition analysis

Separated oligonucleotides isolated from PEI-cellulose were digested to completion using nuclease P₁ or RNAase T₂, and the products were analyzed by high voltage paper electrophoresis on Whatman 3MM paper as described previously (8). Using the autoradiogram as a guide, individual spots were cut from the paper and the radioactivity quantitated by liquid scintillation counting in a toluene-based fluor.

RESULTS

Analysis of 5' end-labeled IVS RNA

Isolated Tetrahymena nuclei were incubated with [³H]UTP under conditions that allow transcription and splicing of the rRNA precursor. The RNA was extracted, and the excised IVS RNA was purified by sucrose gradient centrifugation and preparative gel electrophoresis. This RNA was intact and radiochemically pure, as judged by tritium fluorography of an analytical 8 M urea-polyacrylamide gel (data not shown).

One portion of the IVS RNA was pretreated with bacterial alkaline phosphatase, while another portion was taken through the same procedure but without addition of phosphatase. The samples were then incubated with [γ -³²P]ATP and polynucleotide kinase under pH 9.0 conditions where the forward reaction predominates (13). Six-fold more ³²P was incorporated in the RNA that had been treated with phosphatase than in the control sample. This result provided the first evidence that the IVS RNA contained a 5'-phosphate, removal of which was necessary to achieve efficient labeling with kinase. Most of the end-labeled RNA was intact, as judged by gel electrophoresis (Figure 1A). Southern hybridization of the end-labeled RNA to a Hha I digest of purified Tetrahymena rDNA and to a Hha I digest of an IVS-containing recombinant plasmid (pTT116; ref. 14) confirmed that the label had been incorporated into IVS RNA (data not shown).

The nucleotide sequence of the 5' end-labeled IVS RNA was determined by partial digestion with base-specific RNAases followed by gel electrophoresis (9,15). From the sequencing gel (Figure 1B), 20 of the first 26 nucleotides (beginning with the sequence AAUAG) could be identified. The RNA sequence (Figure 1C) is colinear with the DNA sequence determined by Kan and Gall (14).

From the sequencing gel, it appeared that there was an unidentified nuc-

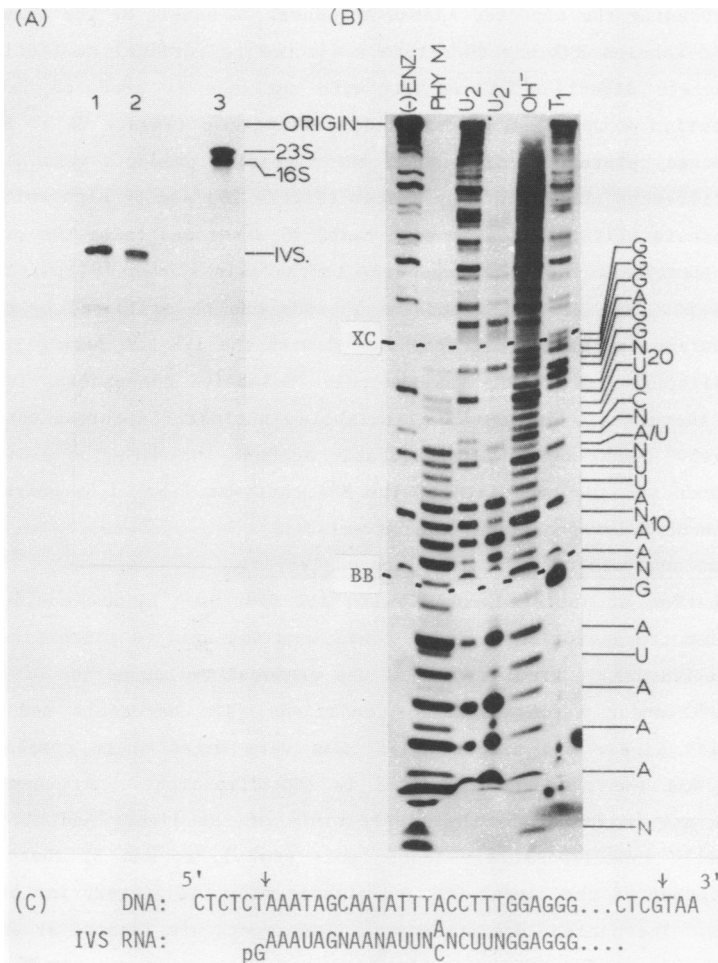


Figure 1. Nucleotide sequence of the IVS.

(A) 4% polyacrylamide, 8 M urea gel showing the purity of the end-labeled linear IVS RNA which was used for sequencing and end analysis. (Lane 1) End-labeled IVS RNA. (Lane 2) Uniformly ³²P-labeled IVS RNA. (Lane 3) *E. coli* rRNA markers.

(B) RNA sequencing gel. Purified end-labeled IVS shown in (A) was subjected to partial digestion with RNAases Phy M (cleaves after A and U), U₂ (cleaves after A), and T₁ (cleaves after G). Partial alkaline hydrolysis (OH⁻) produced a ladder, marking the position of all the nucleotides. RNA incubated without enzyme ((-)ENZ) was cleaved to some extent, preventing us from specifying some of the nucleotides (designated "N"). Dye markers: XC, xylene cyanole; BB, bromophenol blue. From this data, the nucleotide preceding the first A could not be identified. The 5' terminal nucleotide was later identified as a G (Fig. 2).

(C) The RNA sequence is compared to the sequence of the IVS region of *T. thermophila* rDNA (14). The arrows indicate the splice junctions (14).

leotide preceding the expected AAAUAG sequence. A sample of the same preparation of end-labeled RNA was therefore subjected to terminal nucleotide analysis. Complete digestion of the RNA with nuclease P₁ produced labeled pG, while digestion with RNAase T₂ produced labeled pGp (Figure 2A). Because of the unexpected nature of this result, the digestion products were also analyzed by a different chromatography system (Figure 2B) and by high voltage paper electrophoresis (Figure 2C). In each case, pG was found to be the predominant 5' end. Quantitation of the paper electrophoresis showed 79% pG, 11% pA, 6% pU, and 4% pC. The 21% non-guanosine 5' ends can be explained by contaminating RNA molecules that co-electrophoresed with the IVS RNA during its preparation. (Although the IVS RNA was the only ³H-labeled RNA species in the preparation, there was a low level of unlabeled nuclear RNA throughout the preparative gel. Such contaminants probably account for both the 21% non-guanosine 5' ends and the one-sixth of the RNA that was labeled by polynucleotide kinase without prior phosphatase treatment.)

Fingerprint analysis of uniformly labeled IVS RNA

Incubation of isolated nuclei with all four [α -³²P] nucleoside triphosphates under transcription-splicing conditions was used to synthesize uniformly labeled IVS RNA. Circular IVS RNA was produced by incubation of linear IVS RNA at 39°C under autocyclization conditions (see Materials and Methods). Gel-purified linear and circular IVS RNAs were digested to completion with RNAase T₁ and the products separated in two dimensions. As seen from the autoradiograms in Figure 3, the fingerprints of the linear and circular RNAs are virtually superimposable. Three spots (no. 1, 2, and 8) that appear in the fingerprint of the linear IVS are missing from the fingerprint of the circular form. The nucleotides present in these spots are located at or near the ends of the linear IVS RNA (see below).

The base composition of each oligonucleotide in the linear IVS RNA fingerprint was determined by considering the location of the spot on the fingerprint (11) and by secondary digestions with RNAase T₂ and nuclease P₁ (e.g., Figure 4). The first residue of each oligonucleotide, which loses its label upon nuclease P₁ digestion, was in some cases identified as the nucleotide that appeared or increased in amount in the RNAase T₂ digestion. The sizes of some of the oligonucleotides were confirmed by electrophoresis in a 20% polyacrylamide gel. Finally, the base sequence of each oligonucleotide was specified by reference to the DNA sequence (14). The base composition and identification of the oligonucleotides are given in Table 1.

Only the central portion of the first dimension cellulose acetate strip

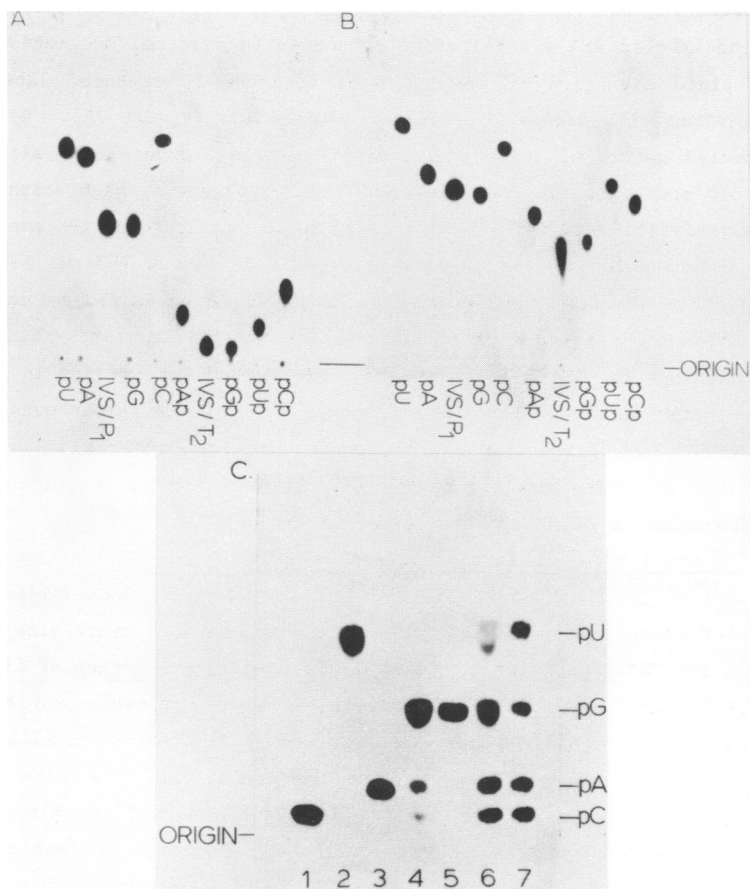


Figure 2. The linear IVS RNA has a 5' terminal guanosine. (A) Purified IVS, labeled at its 5' end with ^{32}P , was totally digested with either nuclease P_1 or RNAase T_2 , and analyzed by PEI-cellulose ascending chromatography using 1 M LiCl (pH 7.0) as the solvent. Nuclease P_1 released a labeled pC, and RNAase T_2 released a labeled pGp. (B) Same samples as in (A) were chromatographed in 1 M sodium formate (pH 3.5) solvent. (C) High voltage paper electrophoresis in pyridine-acetic acid buffer (pH 3.5). (Lane 1) pC. (Lane 2) pU. (Lane 3) pA. (Lane 4) IVS RNA digested with nuclease P_1 . (Lane 5) pG. (Lane 6) Uniformly labeled 23S *E. coli* rRNA digested with nuclease P_1 . (Lane 7) Mixture of the four pN markers.

was blotted onto PEI-cellulose for separation in the second dimension. Autoradiography of the remaining portions of the cellulose acetate strip revealed no radioactivity at a sensitivity where one phosphate per RNA molecule was readily detected. The absence of radioactivity near the origin provides an argument against the RNA containing a covalently bound protein of the type

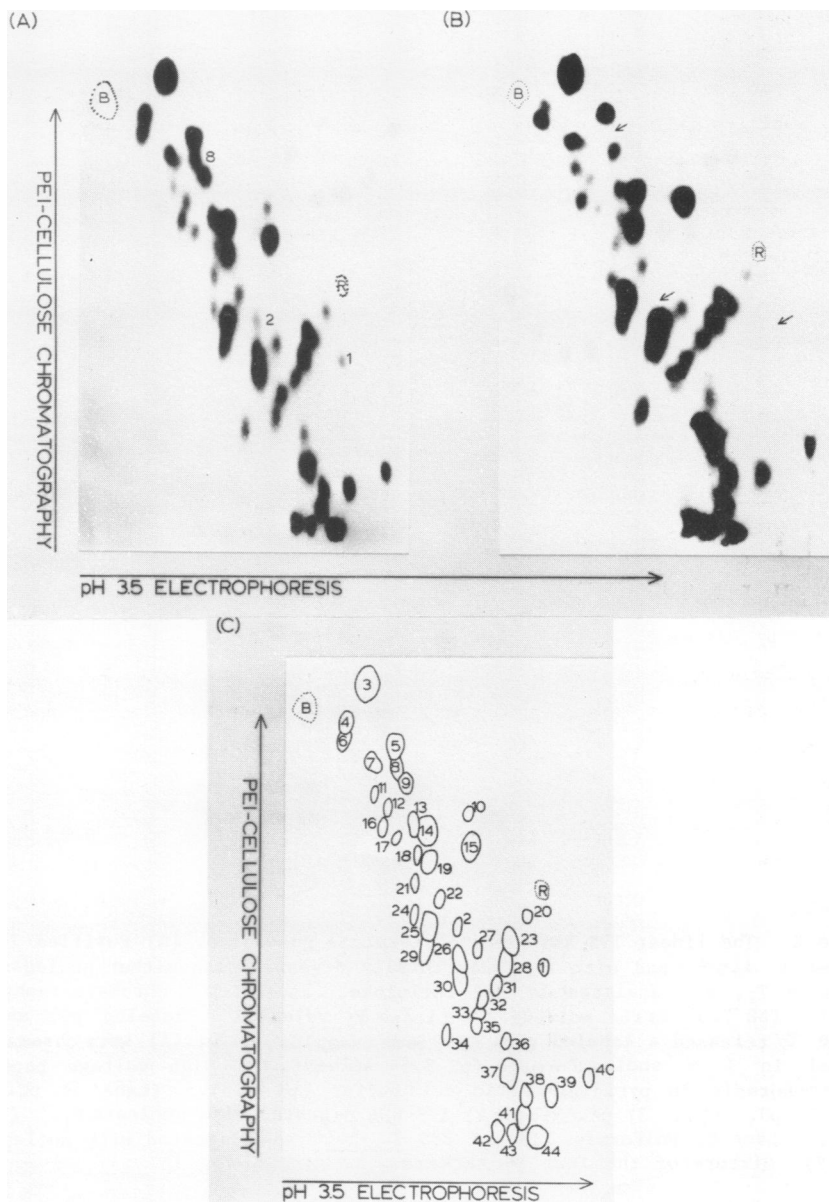


Figure 3. Fingerprint analysis of uniformly labeled IVS RNA. (A) Linear IVS RNA was digested to completion with RNAase T₁ and the products were separated in two dimensions as described in Materials and Methods. Positions of dye markers are indicated (B) xylene cyanole, (R) acid fuchsin. (B) Circular IVS RNA analyzed as described in (A). Arrows indicate the positions of T₁ oligonucleotides which were present in the linear IVS RNA but missing in the circular form. (C) Diagrammatic representation of linear IVS RNA fingerprint (A).

found on the 5' terminus of poliovirion RNA (16,17). If it had been found, such a protein could have provided an explanation for the autocyclization activity of the IVS RNA. The absence of radioactivity with higher electrophoretic mobility than pGp eliminates the possibility of any significant amount of di- or tri-phosphorylated 5' termini.

The 5' end. Spot no. 1 on the RNA fingerprint of Figure 3A contained pGp, which must arise from the 5' end of the linear molecule. Its molar yield was determined to be 2.1 (Table 1). Considering the 2-fold overrepresentation of other small products (e.g., Gp, AGp and UGp) in the fingerprint, the corrected molar yield of pGp would be 1.0. The results are therefore consistent with the IVS RNA having a homogeneous pGp terminus. Spot no. 2 contained AAAUAGp, which follows the initial guanosine in the sequence. Both spots are completely missing from the fingerprint of the circular IVS RNA (Figure 3B).

The 3' end. Spot no. 8 had a base composition consistent with it being a hexanucleotide (Table 1), yet it chromatographed on PEI-cellulose with the trinucleotides (Figure 4). This chromatographic behavior is consistent with the oligonucleotide having a 3'-hydroxyl group rather than the 3'-phosphate produced by RNAase T₁ cleavage. Digestion of the RNA eluted from spot no. 8 with nuclease P₁ liberated only 0.2 moles of inorganic phosphate (Table 2), while the other spots in the fingerprint contained 1.0 ± 0.3 (mean \pm SD) moles of terminal phosphate. Digestion with RNAase T₂ liberated only 0.2 moles of Gp. From these results it was concluded that 80% of spot 8 was the 3'-terminal hexanucleotide UACUCG·OH, while 20% was a contaminating oligonucleotide ending in Gp. In a separate experiment, spot 8 RNA was treated with phosphatase and then electrophoresed on a 20% sequencing gel. The electrophoretic mobility of the predominant oligonucleotide did not change upon phosphatase treatment, which is the result expected for a T₁ oligonucleotide with a 3'-hydroxyl. These results do not provide a direct demonstration of a 3'-hydroxyl, but do allow us to exclude the possibility of a phosphate on the 3' or 2' position of the terminal guanosine (see Discussion). (A minor component of spot no. 8, a dinucleotide, did show a great reduction in electrophoretic mobility upon phosphatase treatment, indicating removal of a 3'-phosphate. This dinucleotide apparently accounted for the small amount of inorganic phosphate liberated from spot 8 RNA by nuclease P₁ treatment, as described above.) The molar yield of spot 8 RNA was near unity, consistent with the linear IVS RNA having a homogeneous 3' end. Spot 8 is completely missing from the fingerprint of the circular IVS RNA (Figure 3B).

Nucleic Acids Research

Table 1. Identification of oligonucleotides (oligos) produced by RNAase T₁ digestion of linear IVS RNA.

Spot no.	Molar yield		Base composition observed	Sequence
	Observed	Expected		
1	2.1	1.0	pGp	pGp
2	1.1	1.0	N(2.2A, 0.0C, 0.7U)Gp	AAAUAG
3	68.0	34.0	Gp	G
4	2.2	3.0	C(0.0A, 0.0C, 0.0U)Gp	CG
5	10.0	5.0	A(0.0A, 0.0C, 0.0U)Gp	AG
6	2.1	2.0	C(0.0A, 1.0C, 0.0U)Gp	CCG
7	2.6	3.0	N(1.1A, 0.3C, 0.0U)Gp	ACG, CAG(2)
8	1.7	1.0	N(0.6A, 1.6C, 0.5U)G•OH	UACUCG
9	3.2	2.0	A(1.4A, 0.0C, 0.0U)Gp	AAG
10	2.1	1.0	U(0.0A, 0.0C, 0.0U)Gp	UG
11	0.9	1.0	A(0.0A, 2.1C, 0.0U)Gp	ACCG
12	0.8	1.0	C(1.1A, 0.0C, 0.0U)Gp	CAAG
13	2.1	2.0	A(2.4A, 0.0C, 0.0U)Gp	AAAG
14	6.0	5.0	N(0.0A, 0.8C, 0.7U)Gp	CUG(2), UCG(3)
15	9.1	6.0	N(1.0A, 0.0C, 0.6U)Gp	AUG(5), UAG(1)
16	1.0	1.0	N(2.8A, 0.7C, 0.0U)Gp	CCAAG
17	0.9	1.0	C(4.8A, 0.0C, 0.0U)Gp	CAAAG
18	1.3	1.0	A(3.0A, 0.4C, 0.4U)Gp	AAAAG
19	5.4	3.0	C(1.7A, 0.3C, 1.1U)Gp	CUAG(2), CAUG
20	1.0	1.0	U(0.0A, 0.0C, 1.9U)Gp	UUG
21	1.3	1.0	N(1.4A, 0.9C, 0.8U)Gp	CAUCG
22	0.8	1.0	N(1.1A, 0.8C, 0.8U)Gp	ACAUG
23	2.9	3.0	N(1.0A, 0.0C, 1.7U)Gp	AUUG, UAUG(2)
24	0.4	1.0	N(2.0A, 2.2C, 0.6U)Gp	CAACACUG
25	1.4	2.0	N(4.0A, 2.5C, 1.1U)Gp	UACCAAG, UCAACAG
26	1.9	2.0	N(1.1A, 1.8C, 1.6U)Gp	UCUCAG, CCUUG
27	1.2	1.0	N(1.1A, 1.1C, 1.2U)Gp	UUCAG
28	2.3	2.0	Not determined	AUAUG, AUUAG
29	0.8	1.0	N(4.0A, 2.6C, 0.8U)Gp	UACCAAG
30	1.3	2.0	N(1.1A, 1.1C, 0.6U)Gp	UCCUAAG, UUCACAG
31	0.9	1.0	N(1.6A, 0.0C, 2.0U)Gp	AUAUAG
32	0.4	1.0	N(2.0A, 0.0C, 0.6U)Gp	UAAUAAG
33	0.7	1.0	N(2.8A, 1.1C, 1.6U)Gp	ACUAAAUG
34	0.5	1.0	N(3.6A, 5.3C, 0.8U)Gp	UCCUAACCAG

Spot no.	Molar yield		Base composition observed	Sequence
	Observed	Expected		
35	0.7	2.0	N(3.4A, 1.4C, 2.3U)Gp	UCAAUUG, AAACUUUG
36	0.7	1.0	N(2.5A, 1.2C, 2.6U)Gp	UUAUCAG
37	-	1.0	Not determined	UUUAAAAG
38	1.4	1.0	A(0.8A, 1.8C, 3.5U)Gp	AUCUUCUG
39	1.1	1.0	N(3.0A, 0.1C, 4.1U)Gp	UAUAUUG
40	1.5	1.0	U(0.0A, 0.0C, 3.9U)Gp	UUUUG
41	1.2	1.0	N(3.7A, 1.6C, 4.5U)Gp	AACUAAUUUG
42	0.5	1.0	N(3.8A, 7.7C, 7.2U)Gp	ACCUCUCCUAAUG
43	0.4	1.0	Not determined	UCUUUAAAACCAAUG
44	1.5	2.0	N(6.6A, 3.9C, 9.4U)Gp	UAUUCUUCUCAUAG, CAAUAUUUACCUUG

Observed molar yields were calculated as $(R_n/P_n)/(R/P)$, where R_n is the radioactivity in oligo n , P_n is the number of phosphates in the oligo, R is the total radioactivity in all spots, and P is the total number of phosphates (414 in the entire IVS; ref. 14). Values obtained from 2-3 different RNA preparations are averaged. Expected molar yields give the number of times that the oligo occurs in the DNA sequence (14). Discrepancies between observed and expected molar yields were due to unequal incorporation of the four nucleoside triphosphates during transcription, which varied in an unpredictable fashion from experiment to experiment. In addition, the transfer of oligos from the cellulose acetate strip to the PEI-cellulose plate and the elution of oligos from the PEI-cellulose were not totally complete. These problems appeared to be more severe for longer oligos. The observed molar yields are therefore systematically overestimated for short oligos and underestimated for long oligos. Observed base compositions: The molar ratio of each base relative to 1.0 mole of G is given in parentheses. Ratios were determined by digestion with nuclease P_1 , which leaves each labeled phosphate attached to the nucleoside with which it was incorporated. [This allowed correction for unequal incorporation of the four nucleoside triphosphates during transcription; the correction factors were determined by P_1 digestion of whole IVS RNA, using the knowledge that the IVS contains 123 A's, 76 C's, 107 U's and 108 G's (14).] The first residue of each oligo loses its label upon nuclease P_1 digestion and therefore does not contribute to the base ratios in parentheses. In some cases this residue was identified as the nucleotide that appeared or increased in amount upon RNAase T_2 digestion of the oligo. In other cases the RNAase T_2 results were inconclusive, so the first residue is designated as "N". Oligos were assigned to the DNA sequence (14) as described in the text. The identifications are tentative for spots 26 (which contained a pentanucleotide and a hexanucleotide), 25, and 29. One oligonucleotide predicted by the DNA sequence, CACCUG, was not observed in the fingerprints. An oligo with that base composition would be located to the left of spot 21.

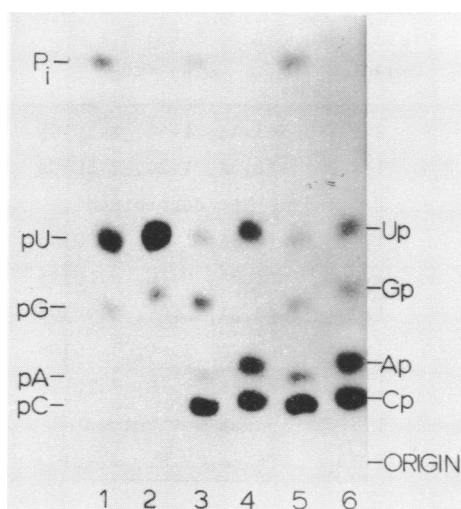


Figure 4. Secondary digests of isolated oligonucleotides. RNAase T₁ oligonucleotides were recovered from fingerprints (Fig. 3A) and totally digested with nuclease P₁ (lanes 1,3,5) or RNAase T₂ (lanes 2,4,6). The products were separated by high voltage paper electrophoresis. (Lanes 1 and 2) Spot 40. (Lanes 3 and 4) Spot 8. (Lanes 5 and 6) Spot 34. The low yield of P₁ after P₁ digestion (lane 3) and the disappearance of labeled G upon T₂ digestion (lane 4) indicate that there is a 3'-terminal G·OH on oligonucleotide no. 8. The apparent underrepresentation of pA in lanes 3 and 5 is described in the legend to Table 2.

Table 2. Evidence that the IVS RNA has a 3' terminal G·OH.

Lane no.	1		3		5	
	Spot no. 40		Spot no. 8		Spot no. 34	
	Base composition					
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
P ₁	1.1	1	0.2	0	1.0	1
pU	2.7	3	0.7	1	0.7	1
pG	1.0	1	1.0	1	1.0	1
pA	0	0	0.8*	1	3.0*	3
pC	0	0	1.9	2	4.4	5

Lane numbers correspond to Figure 4. Observed (Obs.) base compositions of oligonucleotides were determined by nuclease P₁ digestion. Expected (Exp.) values are based on the identifications given in Table 1. Note that spot 8, the 3' terminal oligonucleotide, has a very low ratio of P₁/pG compared to the other oligonucleotides. *Values for pA were corrected as described in the legend to Table 1 to account for a 2.8-fold low incorporation of this nucleotide relative to the other three during transcription.

DISCUSSION**Structure of the linear IVS RNA**

We have shown that the linear form of the excised IVS RNA begins with a guanosine residue not predicted by the DNA sequence of the IVS (14). The guanosine residue was found both by 5'-terminal nucleotide analysis of end-labeled IVS RNA and by RNAase T₁ fingerprinting of uniformly labeled IVS RNA. It initially seemed possible that the guanosine might be transferred from the 3' end of the IVS RNA during splicing. We can now exclude this possibility, because the 3'-terminal oligonucleotide (spot no. 8, Table 1) contains a stoichiometric quantity of guanosine as revealed by digestion with nuclease P₁ (lane 3 of Figure 4 and corresponding part of Table 2).

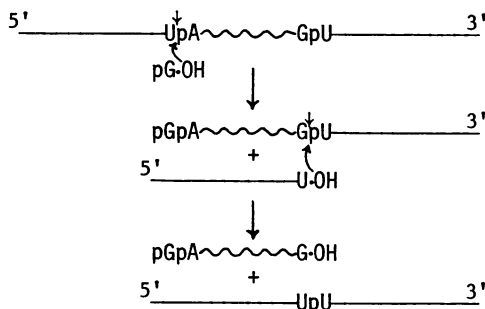
More recently, we have found that a guanosine residue is added to the 5' end of the IVS RNA during splicing. We have purified a form of unspliced pre-rRNA which is capable of excising its IVS when incubated with a guanosine compound in the presence of monovalent and divalent cations (8). In the cases of GTP and GMP, we showed that the guanosine moiety becomes covalently attached to the 5' end of the IVS RNA via a normal phosphodiester bond (8). In this highly purified in vitro splicing system, GTP, GDP, GMP and guanosine were found to be approximately equally effective as cofactors in the splicing reaction (8). In contrast, the IVS RNA excised from the pre-rRNA in isolated nuclei has an apparently homogeneous monophosphoryl 5' terminus (this work). [The fingerprinting data preclude di- and tri-phosphate 5' ends. The low amount of labeling of the IVS by polynucleotide kinase in the absence of phosphatase treatment and the high molar yield of pGp in the fingerprint preclude a major fraction of hydroxyl 5' ends.] Therefore, in isolated nuclei (and presumably in vivo) a specific guanosine compound may serve as the cofactor for splicing. GTP could be used as the cofactor, followed by cleavage of inorganic pyrophosphate. Alternatively, GMP could be used directly.

The IVS RNA has a 3'-terminal guanosine residue, as predicted from the DNA sequence (14). The data of Figure 4 preclude this guanosine having a 2' or 3' phosphate, and are consistent with it having free 2' and 3' hydroxyl groups. We have not performed a direct test for free hydroxyl groups. Some modifications such as 2'-O-methylation and base methylation do not produce much of a change in a nucleotide's migration in high voltage paper electrophoresis (11), and might escape notice. Methylation is not expected, however, because the nuclear transcription system used to synthesize the IVS does not contain a methyl donor such as S-adenosylmethionine or its precursor, methionine. We tentatively conclude that the 3' end of the RNA is an unmodified

G•OH.

By what pathway is the IVS excised with a 5'-terminal pGpA... and an unphosphorylated 3' end? Such termini would be produced by endonucleolytic cleavage of the pre-rRNA at the two splice junctions, generating 5'-pA and 3'-G•OH termini on the IVS, followed by addition of the guanosine cofactor (which must have a free 3'-hydroxyl; ref. 8) to the 5' end of the molecule. The production of 5'-phosphoryl and 3'-hydroxyl termini is normal for endonucleases involved in RNA processing (1). In the splicing of yeast tRNA precursors, however, the endonuclease step produces 5'-hydroxyl and 3'-phosphoryl termini (4). Therefore, in the two types of RNA splicing systems for which such information is available, the endonucleolytic steps are fundamentally different.

In the case of yeast tRNA splicing, Peebles *et al.* (18) obtained evidence for an ATP-dependent ligation activity that joined the termini of the tRNA exons to generate a mature tRNA molecule. We have not been able to isolate the free exons of the *Tetrahymena* pre-rRNA, and have no direct evidence concerning the ligation step. It is possible that rRNA splicing, like tRNA splicing, is a two-stage cleavage and ligation reaction (18). It is easier to rationalize the addition of the guanosine nucleotide and the lack of an ATP requirement, however, by postulating that cleavage and rejoining occur in a concerted fashion (8). This phosphoester transfer mechanism is depicted below, with the IVS indicated by a wavy line and the exons by straight lines.



Note that the cleavage steps, indicated by the short vertical arrows, produce 5'-phosphoryl and 3'-hydroxyl groups on all reaction products, and that the transfer steps involve the joining of the 5'-phosphate to a new 3'-hydroxyl. Concerted cleavage-ligation events involving such termini are catalyzed by the type I and II DNA topoisomerases of prokaryotes (19). In the case of the topoisomerases, however, the 5'-phosphate is rejoined to the 3'-hydroxyl group

of the same nucleotide to which it was originally attached.

In their DNA sequencing study, Kan and Gall (14) found that one cloned copy of the 1.6 kilobase Hind III fragment of rDNA had the same 413 base pair IVS sequence found in genomic rDNA, while another cloned copy had a variant sequence 7 base pairs longer. IVS RNA produced from the genomic rDNA sequence is predicted to give a unique T₁ oligonucleotide, CAAAG. This oligonucleotide was clearly present in the IVS RNA (spot no. 17, molar yield = 0.9), confirming that the 413 pair sequence was correctly assigned as the major genomic version. The variant rDNA sequence is predicted to give two unique T₁ oligonucleotides, CUUG and AAUG, neither of which was present in the IVS RNA. Therefore, genes with the variant IVS sequence, if they occur at all in genomic rDNA (see ref. 14), do not produce a significant portion (>10%) of splicable pre-rRNA.

Structure of the circular IVS RNA

Three RNAase T₁ oligonucleotides present in fingerprints of linear IVS RNA were completely absent in fingerprints of circular IVS RNA. They were identified as the 5'-terminal pGp (spot 1), the hexanucleotide that follows the terminal guanosine (spot 2), and the 3'-terminal hexanucleotide (spot 8). If the circle were produced by simple end-to-end ligation of the linear IVS RNA, the junction would have the sequence 5'-UpApCpUpCpGpGpApApUpApGp-3', where the arrow indicates the point of ligation. RNAase T₁ digestion would then produce no spot 1 or spot 8 RNA, but the spot 2 oligonucleotide would be unaltered. The disappearance of spot 2 leads us to suggest that cyclization involves attack of the 3'-terminal hydroxyl group of the IVS on a phosphate near but not at the 5' end of the molecule. This view is supported by another observation: when 5' end-labeled IVS is cyclized, no radioactivity is found in the circular RNA product (8). We have not yet found any new oligonucleotide in the fingerprint of the circular IVS RNA which represents the cyclization junction. This has hindered the determination of the primary structure of the circular IVS RNA. Based on the changes in the RNA fingerprint, it is clear at this time that IVS RNA cyclization does involve cleavage and reformation of covalent bonds.

ACKNOWLEDGMENTS

We are grateful to Gary Olsen for help in RNA sequencing and to Linda Breeden and Mike Yarus for help in fingerprinting. We thank Nancy Kan and Joe Gall for sharing the DNA sequence of the IVS prior to its publication. Special thanks to Elaine Black for careful preparation of the manuscript. This

work was supported by grants no. GM28039 and GM25273 from the NIH. T.R.C. was supported by NIH Research Career Development Award no. CA00700.

REFERENCES

1. Abelson, J. (1979) Ann. Rev. Biochem. 48, 1035-1069.
2. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
3. Cech, T.R., Zaug, A.J., Grabowski, P.J. and Brehm, S.L. (1982) in The Cell Nucleus, eds. Busch, H., and Rothblum, L., vol. X, Academic Press, New York, in press.
4. Knapp, G., Ogden, R.C., Peebles, C.L. and Abelson, J. (1979) Cell 18, 37-45.
5. Zaug, A.J. and Cech, T.R. (1980) Cell 19, 331-338.
6. Carin, M., Jensen, B.F., Jentsch, K.D., Leer, J.C., Nielsen, O.F. and Westergaard, O. (1980) Nucl. Acids Res. 8, 5551-5566.
7. Grabowski, P.J., Zaug, A.J. and Cech, T.R. (1981) Cell 23, 467-476.
8. Cech, T.R., Zaug, A.J. and Grabowski, P.J. (1981) Cell 27, 487-496.
9. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Nucl. Acids Res. 4, 2527-2538.
10. Donis-Keller, H. (1980) Nucl. Acids Res. 8, 3133-3142.
11. Barrell, B.G. (1971) Procedures in Nucl. Acid Res. 2, 751-779.
12. Griffin, B.E. (1971) FEBS Lett. 15, 165-168.
13. van de Sande, J.H., Kleppe, K. and Khorana, H.G. (1973) Biochemistry 12, 5050-5055.
14. Kan, N.C. and Gall, J.G. (1982) adjoining paper.
15. Simoncsits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Guillely, H. (1977) Nature 269, 833-836.
16. Lee, Y.F., Nomoto, A., Detjen, B.M. and Wimmer, E. (1977) Proc. Natl. Acad. Sci. USA 74, 59-63.
17. Flanagan, J.B., Pettersson, R.F., Ambros, V., Hewlett, M.J. and Baltimore, D. (1977) Proc. Natl. Acad. Sci. USA 74, 961-965.
18. Peebles, C.L., Ogden, R.C., Knapp, G. and Abelson, J. (1979) Cell 18, 27-35.
19. Gellert, M. (1981) Ann. Rev. Biochem. 50, 879-910.