MOLECULAR AND CELLULAR BIOLOGY, Sept. 1989, p. 3657–3666
0270-7306/89/093657-10$02.00/0

Vol. 9, No. 9

The Conserved U·G Pair in the 5′ Splice Site Duplex of a Group I Intron Is Required in the First but Not the Second Step of Self-Splicing

ELISABETH T. BARFOD† AND THOMAS R. CECH*

Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215

Received 7 April 1989/Accepted 30 May 1989

Group I self-splicing introns have a 5′ splice site duplex (P1) that contains a single conserved base pair (U·G). The U is the last nucleotide of the 5′ exon, and the G is part of the internal guide sequence within the intron. Using site-specific mutagenesis and analysis of the rate and accuracy of splicing of the *Tetrahymena thermophila* group I intron, we found that both the U and the G of the U·G pair are important for the first step of self-splicing (attack of GTP at the 5′ splice site). Mutation of the U to a purine activated cryptic 5′ splice sites in which a U·G pair was restored; this result emphasizes the preference for a U·G at the splice site. Nevertheless, some splicing persisted at the normal site after introduction of a purine, suggesting that position within the P1 helix is another determinant of 5′ splice site choice. When the U was changed to a C, the accuracy of splicing was not affected, but the K_m for GTP was increased by a factor of 15 and the catalytic rate constant was decreased by a factor of 7. Substitution of U·A, U·U, G·G, or A·G for the conserved U·G decreased the rate of splicing by an even greater amount. In contrast, mutation of the conserved G enhanced the second step of splicing, as evidenced by a *trans*-splicing assay. Furthermore, a free 5′ exon ending in A or C instead of the conserved U underwent efficient ligation. Thus, unlike the remainder of the P1 helix, which functions in both the first and second steps of self-splicing, the conserved U·G appears to be important only for the first step.

Group I introns fold into a specific structure that allows them to accomplish precise splicing of flanking RNA sequences (3, 9). Splicing occurs through two transesterification reactions, exchanges of phosphate esters in which the total number of O—P bonds remains constant (Fig. 1A). In the first step, a free molecule of guanosine or GTP breaks the RNA at the 5′ splice site and becomes covalently attached to the 5′ end of the intron (2, 10). In the second step, the newly released 5′ exon attacks the 3′ splice site, simultaneously ligating the exons and excising the intron.

The mechanism by which the 5′ splice site is specified has been partially established. Davies et al. (13) first proposed that the last few nucleotides of the 5′ exon were recognized by base pairing to an internal guide sequence (IGS) within the intron. This interaction was also apparent in the structure models of Michel et al. (22). The role of this base-pairing interaction in specifying the 5′ splice site was proven for the case of the *Tetrahymena* group I intron by analysis of single-base changes and second-site suppressors (5, 31). The same base-pairing interaction contributes to the second step of self-splicing (5, 16, 31) (Fig. 1A).

Only one of the base pairs of P1 is conserved in sequence among diverse group I introns. The nucleotide preceding the 5′ splice site is almost always a U and is juxtaposed to a G in the IGS. The U is present in 64 of the 66 available group I intron sequences; the other two have a C residue (9). Although it is commonly assumed that the U and G form a U·G wobble base pair (12), there is no evidence (phylogenetic or otherwise) for such pairing, and alternative H-bonding interactions have been considered (17). The site for intervening sequence (IVS) RNA cyclization is also marked by a U·G pair (6; Fig. 1A).

An examination of Fig. 1A makes it seem reasonable to propose that the conserved U and G is important for both the first and second steps of self-splicing, but no direct test of the contribution of these nucleotides to splicing has been reported. We now show that the conserved U·G pair is necessary for efficient GTP addition and also contributes to the specificity of this reaction. The U·G pair is not required for the exon ligation step and, surprisingly, is not even the optimal combination of bases for this reaction.

MATERIALS AND METHODS

Plasmid constructions and site-directed mutagenesis. Two different vectors, in earlier work an M13-derived construction and later a pUC-derived plasmid, were used to make single base changes in the *Tetrahymena thermophila* large subunit (LSU) rRNA gene. Although the properties of the corresponding wild-type (unaltered) pre-rRNAs transcribed from the two templates appeared to be the same, the activity of each mutation was compared with that of the wild type of the parent DNA from which it was derived.

The M13 derivative contains a bacteriophage SP6 promoter upstream of a portion of the *Tetrahymena* gene that includes the intron and short regions of the exons. The SP6 promoter with a 42-base-pair leader sequence was isolated on a *Sal*I-*Hin*dIII DNA fragment of pSPTT1A3 (34) and inserted into M13mp8 at the *Sal*I and *Hin*dIII sites in the polylinker. A *Hin*dIII fragment from pJE457 (23) provided 16 nucleotides (nt) of natural 5′ exon, the entire 413-nt intron, and 36 nt of the natural 3′ exon. This fragment was purified, inserted into the *Hin*dIII site of the M13-SP6 subclone, and checked for proper orientation relative to the SP6 promoter. In vitro transcription of this DNA cut with *Pvu*I yields a

* Corresponding author.
† Present address: Department of Biochemistry, University of California, Berkeley, CA 94720.

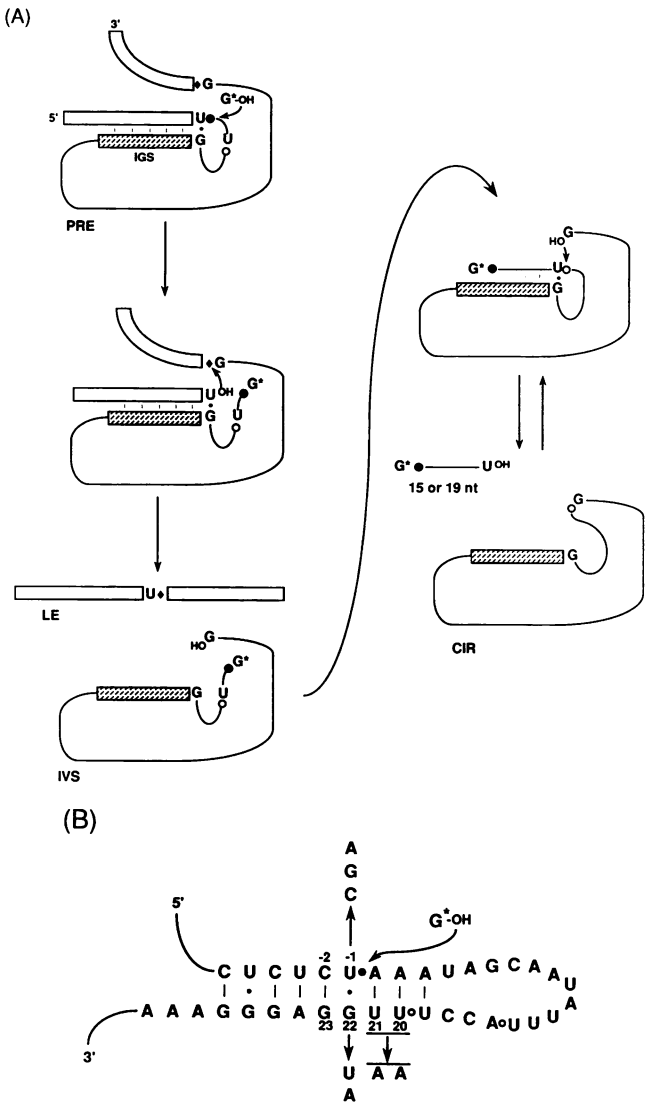


FIG. 1. Involvement of paired region P1 in group I intron reactions. (A) Mechanism of group I intron splicing and cyclization. In precursor RNA containing a group I intron, the 5′ exon forms a duplex P1 with a region internal to the intron (IGS; shaded regions) such that the conserved U at the end of the 5′ exon is aligned opposite a conserved G. In the first step of splicing, a specifically bound GTP substrate cleaves the bond between the 5′ exon and the intron (●) and forms a covalent bond with the intron. The freed 3′ hydroxyl of the 5′ exon then attacks the 3′ splice site (◆) to form ligated exons and liberate the intron. Some group I introns next refold such that an internal U is aligned opposite a G in the IGS (not necessarily the same G involved in splicing). The last nucleotide of the intron, another conserved G, can then attack the phosphate following the U (○) to form a circular molecule and release an oligonucleotide consisting of the nucleotides preceding the cleavage site. The transesterification mechanism and the role of GTP are derived from our work (reviewed by Cech [8]). The contribution of the IGS was first proposed by Davies et al. (13) and elaborated by Waring et al. (31) and Been and Cech (5, 6). (B) Point mutations of the 5′ splice site U·G pair in the *T. thermophila* LSU pre-rRNA. The 5′ splice site is indicated (●). Sequences to the left of this bond (negative numbers) constitute the 5′ exon. Numbering of intron nucleotides begins with 2, the number 1 being reserved for the G added during splicing. ○, Cyclization sites.

precursor of 659 nt with a 5′ exon of 75 nt and a 3′ exon of 171 nt. The point mutations in which the U at the end of the 5′ exon was changed to a purine were generated by site-directed mutagenesis of the M13 plus strand. Oligonucleotides complementary to the single-stranded form, except for a mismatch that codes for the desired mutation, were extended with the Klenow fragment of DNA polymerase. The resulting pieces were then joined with DNA ligase to create a double-stranded plasmid incorporating the mismatch on one strand (37). The mixtures were transformed, screened, plaque purified, and sequenced to give the mutants -1A, -1G, and -1G/AA (-1G:20A:21A, where colons indicate a mutant with multiple base substitutions).

The pUC-derived plasmid pBGST7 contains a phage T7 promoter upstream of the intron-containing sequences and was designed so that accurate splicing of the intron would produce active β-galactosidase in *Escherichia coli* (5). The lengths of the natural *Tetrahymena* exon sequences included in this construction differ from that of the M13 DNA by having 13 fewer nt in the 3′ exon. In vitro transcription of this plasmid cut with *Nar*I yields a precursor of 871 nt with a 5′ exon of 275 nt and a 3′ exon of 183 nt. Mutants -1C, 22A, and 22U were prepared from this parent plasmid. Site-directed mutagenesis again utilized mismatch oligonucleotides and Klenow DNA polymerase, but in this procedure the template was made duplex except in the region of mutagenesis (18). This was achieved by annealing plasmids cut in a selectable marker (*Xmn*I in the ampicillin gene) with an isolated fragment that does not include the gene to be altered (*Eco*RI and *Hin*dIII excise a fragment containing the *Tetrahymena* sequences). Half of the molecules that form a duplex with a strand from each plasmid have a single-stranded region complementary to the mismatched oligonucleotides.

RNA preparation and sequencing. RNA precursors from the T7 constructions were prepared as previously described (1). RNA from the SP6 constructions was prepared similarly except that the DNA was treated with T4 DNA polymerase to remove the 3′ overhang generated by *Pvu*I prior to transcription (27).

Ligated exons were sequenced by primer extension with reverse transcriptase in the presence of dideoxynucleotides (26) as described by Zaug et al. (36) except that 48°C was used to reduce the effect of secondary structure. Gel-purified 5′-^{32}P-labeled species were sequenced by partial RNase digestion (14).

RNA splicing. Reactions were performed essentially as described by Barfod and Cech (1). Uniformly ^{32}P-labeled precursors were preincubated in splicing buffer [100 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 30 mM Tris, pH 7.5] at 30°C for 2 min. Reactions were started by addition of guanosine or GTP. Equal portions were removed at different times, and the reactions were stopped by addition of 25 mM EDTA and formamide on ice. Samples were analyzed on a denaturing 4% polyacrylamide gel.

Kinetic analysis. Polyacrylamide gels were dried and exposed to X-ray film, and all visible bands were excised. The gel slices were counted with fluor in a liquid scintillation counter to quantify the amount of RNA present. The fraction unreacted was calculated as the amount of precursor remaining divided by the sum of precursor plus products. The natural log of the fraction unreacted was plotted as a function of time. Linear regression analysis was applied to datum points that represented less than 20% reaction. The slopes of these lines were taken to be $-k_{obs}$. Hanes plots

([guanosine]/k_{obs} as a function of [guanosine]) were analyzed to determine K_m and catalytic rate constant (k_{cat}).

RESULTS

Conserved U·G is important for 5′ splice site activity. Site-specific mutagenesis was used to introduce single-base substitutions into the conserved U·G pair in P1 of the *Tetrahymena* intron. U(−1) was changed to each of the other three nucleotides, generating mutants -1C, -1A, and -1G (Fig. 1B). G(22) was changed to either a U (mutant 22U) or an A (mutant 22A).

In addition, mutations were designed to disrupt the putative helical extension in which A(2), A(3), and A(4) pair to U(19), U(20), and U(21). It was known from previous work that this interaction was not required for the GTP addition reaction (34). Thus, in the hope of observing some effect of these nucleotides in an already destabilized background, a triple mutant was made in which U(20) and U(21) were changed to A's at the same time that U(−1) was changed to a G. This triple mutant is designated -1G/AA.

RNA was transcribed from each of the mutant templates, purified, and tested for self-splicing. All mutants showed reduced splicing activity. Nevertheless, when incubated for long periods of time (30 min) at high guanosine concentration, each of the mutant RNAs was converted to the same products as was the wild-type RNA (data not shown). The major products were the linear and circular forms of the IVS and the ligated exons. In addition, free 5′ exon, free 3′ exon, and 5′ exon-IVS were produced in a reaction that was largely GTP independent, and therefore ascribed to specific hydrolysis at the 3′ splice site and subsequent reactions (17).

The mutant RNAs underwent splicing at a slower rate than did the wild-type RNA (Fig. 2). In addition, they required higher guanosine concentrations. The -1C RNA underwent less splicing than did the wild-type RNA at 0.05 mM guanosine and was substantially less active with the guanosine analogs inosine and 2-aminopurine ribonucleoside (2AP), which have a K_m 100-fold higher than that of guanosine (2) (Fig. 3). The -1A RNA required even more guanosine and was inactive with the other analogs. The -1G and -1G/AA RNAs gave similar data to those shown for the -1A RNA but were even less active at 0.5 mM guanosine.

In addition to providing information about reaction rate, these data provide information about substrate specificity. Mutations at position −1 did not appear to alter specificity for the nucleotide substrate. Guanosine was a much better substrate than either inosine or 2AP in these mutants, as in the wild-type RNA.

A complete kinetic analysis was performed on some of the mutant RNAs (Table 1). All of the mutants showed substantially increased K_ms for guanosine and reduced k_{cat}s. The net result was that the apparent second-order rate constant for reaction of free guanosine and free ribozyme (k_{cat}/K_m) was reduced 100-fold for the -1C mutant, more than 100-fold for 22U and 22A, and 1,000-fold for -1G. Complete kinetic analyses were also done under another ionic condition, with 200 mM $NH_4C_2H_3O_2$ replacing 100 mM $(NH_4)_2SO_4$ (data not shown). Although individual values of K_m and k_{cat} were affected by the different ionic condition, most of the k_{cat}/K_m values were not substantially different. One exception was the -1C mutant, which was only 5- to 10-fold less active than the wild-type RNA in $NH_4C_2H_3O_2$. The -1A RNA had a k_{cat}/K_m ratio slightly higher than that of the -1G RNA, whereas the -1G/AA RNA had a value slightly lower than that of -1G.

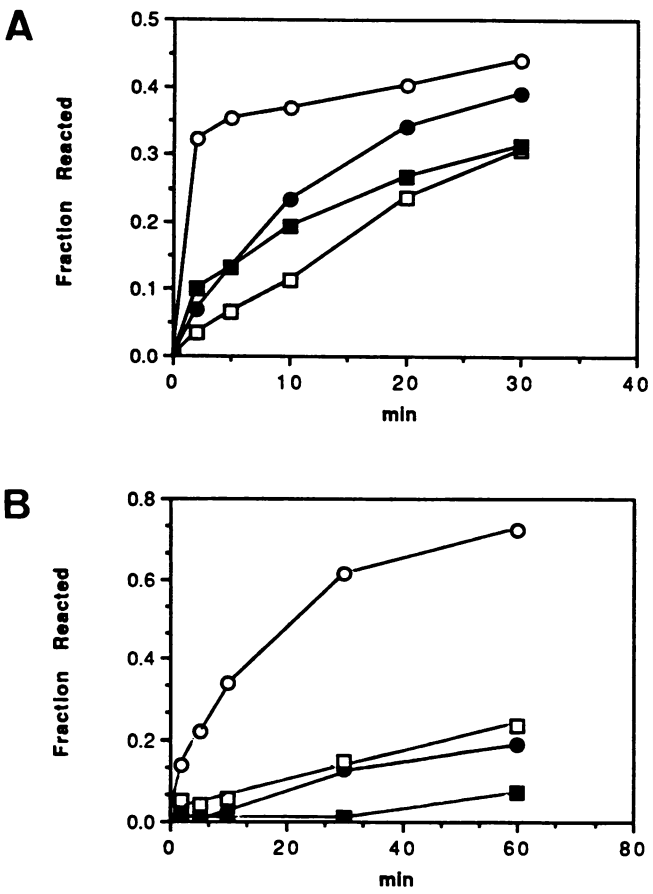
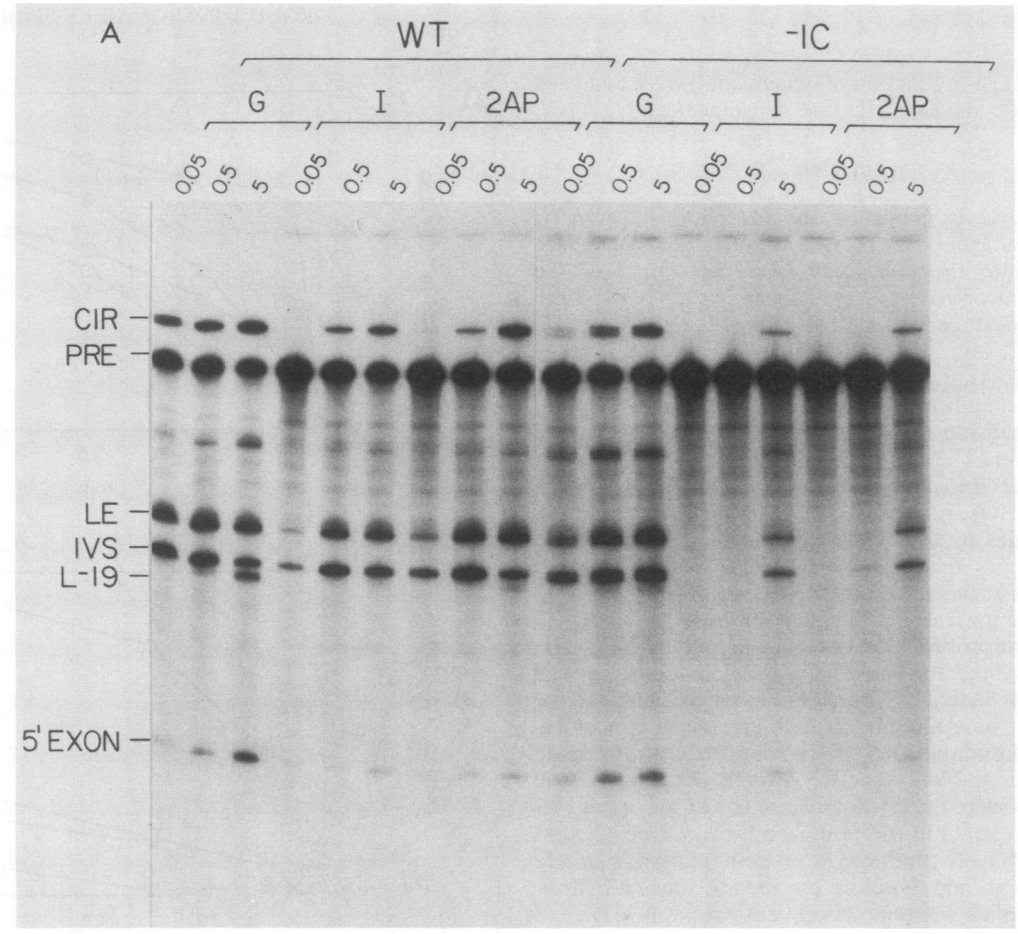


FIG. 2. Reduction of splicing activity by mutation of either nucleotide of the conserved U·G pair. (A) In vitro self-splicing activity of wild-type (○), -1C (●), 22U (□), and 22A (■) precursors. Pre-rRNAs were transcribed with T7 RNA polymerase from *Nar*I-cut plasmid, purified, and renatured by a heat-cool procedure (S. A. Walstrum and O. C. Uhlenbeck, manuscript in preparation). Splicing was under standard conditions with 0.05 mM GTP at 30°C. (B) In vitro self-splicing activity of wild-type (○), -1G (●), -1A (□), and -1G/AA (■) precursors. Pre-rRNAs were prepared with SP6 polymerase from *Pvu*I-cut, filled-in double-stranded M13 DNA. RNA purification and splicing were as described above except that the pre-rRNAs were not heat-cool-treated before the reaction. In both panels, fraction reacted includes products of both splicing and site-specific hydrolysis.

Under conditions of higher temperature (42°C) and higher Mg^{2+} concentration (10 mM), all of the mutants showed increased splicing activity. This activation was primarily due to the higher temperature rather than the increase in Mg^{2+} concentration. Changing the Mg^{2+} concentration from 5 to 10 mM had little effect at either 30 or 42°C.

Activation of a cryptic 5′ splice site. To determine whether the mutations had altered the position of guanosine addition, the sites were mapped. Splicing of ^{3}H-labeled precursor RNAs was carried out with [α-^{32}P]GTP as the substrate. This resulted in 5′ end labeling of the excised IVS and, as cyclization proceeded, release of a 5′-end-labeled oligonucleotide (Fig. 1A).

The -1C, 22U, and 22A mutants generated the same 15- and 19-nt fragments as did the wild-type RNA (Fig. 4A). These two products were gel purified and analyzed by direct enzymatic sequencing (Fig. 5 and data not shown). The sites


A
WT
-IC
G
I
2AP
G
I
2AP
0.05
0.5
5
CIR
PRE
LE
IVS
L-19
5' EXON

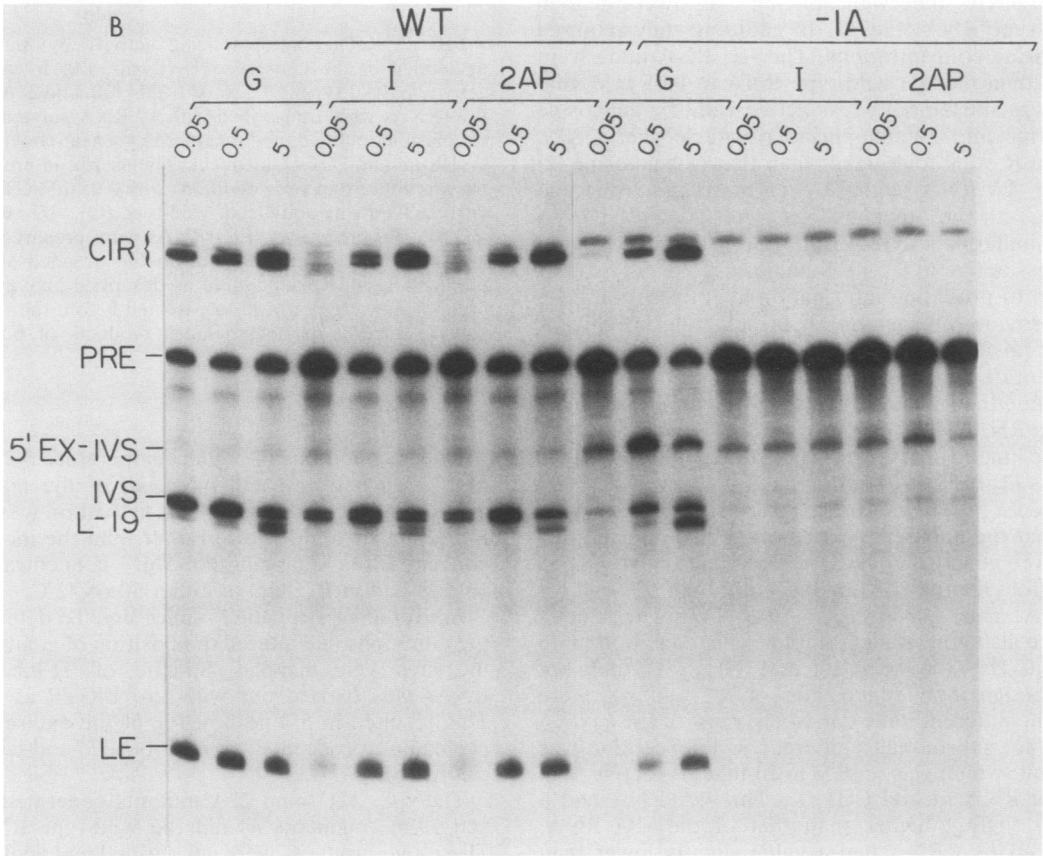

B
WT
-IA
G
I
2AP
G
I
2AP
0.05
0.5
5
CIR
PRE
5' EX-IVS
IVS
L-19
LE

TABLE 1. Kinetic constants for splicing of wild-type and mutant RNAs

RNA	K_m (μM) ± SD[a]	k_{cat} (min^{-1}) ± SD	k_{cat}/K_m ($\times 10^2$) (min^{-1} μM^{-1})
WT[b]	44 ± 7	1.9 ± 0.5	4.3
WT[c]	60 ± 8	1.4 ± 0.2	2.3
-1C	668 ± 83	0.27 ± 0.03	0.04
-1G	630 ± 130	0.015 ± 0.004	0.002
22U	710 ± 250	0.08 ± 0.03	0.01
22A	572 ± 92	0.12 ± 0.02	0.02

[a] SD, Standard deviation of the experimental points from the linear least-squares fit on the Hanes plot.
[b] Wild-type pre-rRNA transcribed from *Hin*dIII-cut pBGST7, the parent vector from which the −1C, 22U, and 22A mutants were constructed.
[c] Wild-type *Tetrahymena* pre-rRNA transcribed from *Pvu*I-cut M13 DNA, the parent vector from which the −1G mutant was constructed.

of guanosine addition and of cyclization were identical to those of the wild-type RNA.

In the case of the -1A and -1G mutants, a major 17-nt fragment was produced (Fig. 4B). This product was pppGCG AAAUAGCAAUAUUU in the case of -1G RNA and pppGCAAAAUAGCAAUAUUU in the case of -1A RNA (Fig. 5). The underlined nucleotides are the last two residues of the mutated 5′ exon. Thus, GTP addition occurred 2 nt before the normal 5′ splice site, following the sequence CUCU. Utilization of this cryptic 5′ splice site is easily understood in terms of the IGS model (Fig. 6D and F).

The 15-nt cyclization products of the -1A and -1G RNAs had the same sequence, pppGAAAUAGCAAUAUUU (Fig. 5). Thus, GTP continued to attack at the normal site even when it was not preceded by a U residue. The 19-nt cyclization products showed a mixed sequence. One component extended from the normal 5′ splice site to the minor cyclization site at position 19; the other included the last 4 nt of the mutated 5′ exon and extended to the major cyclization site at position 15. Thus, a small amount of GTP addition occurred at a cryptic splice site after U(−5).

Inaccuracy of exon ligation in some mutants. Ligated exons were isolated and sequenced by the dideoxynucleotide-chain termination method (data not shown). The 22U and 22A ligated exons had the wild-type sequence CUCUCU/UAA; it should be noted that the technique is not sensitive enough to detect a low level of missplicing. The -1C ligated exons showed the similar sequence CUCUCC/UAA, in which the nucleotide preceding the ligation junction is the C of the mutation.

The -1G and -1A ligated exons were identical to those of the wild type through the 3′ exon. After the splice junction, two sequences were superimposed. Junction sequences CU CUCA/UAA and CUCU/UAA were evident in the -1A ligated exons; CUCUCG/UAA and CUCU/UAA were present in the -1G ligated exons. These sequences are those expected from the two major sites of GTP addition described above. The two sequences were of equal intensity in the -1A ligated exons, consistent with the equal amounts of 15- and 17-nt fragments produced by cyclization (Fig. 4B). Similarly, the major -1G ligated exon species had the CUCU/UAA junction, reflecting the bias in GTP addition seen with this mutant in Fig. 4B. Thus, the efficiency of exon ligation paralleled the availability of a free 5′ exon produced by GTP cleavage. Ligated exons could be formed with any nucleotide as the nucleophile attacking the 3′ splice site.

Second step of splicing. The *trans*-splicing reaction allows the exon ligation step to be assayed by itself, independent of the GTP addition step. pre-rRNA is incubated in the absence of GTP with an oligonucleotide that has a sequence similar to that of the end of the 5′ exon. This miniature version of a 5′ exon attacks the 3′ splice site, becoming ligated to the 3′ exon (16). The reaction involves base pairing of the oligonucleotide to the IGS (5).

Pentanucleotides of sequence C_4N (N is C, U, A, or G) were incubated with several of the pre-rRNAs. (C_5 had previously been shown to be a good substrate for this reaction [16].) The 22U and 22A mutant RNAs produced more of the C_4N- 3′ exon *trans*-splicing product than did the wild-type RNA (Fig. 7A). The reaction conditions were the same as those under which the 22U and 22A mutant precursors had greatly reduced reactivity toward GTP addition (Fig. 4). Furthermore, C_5 was more active than C_4U with all precursors, and C_4A was more active than C_4U or C_5 with the 22U precursor (Fig. 7A). The data clearly show that a G at position 22 and a U at the 3′ end of the 5′ exon are both dispensable for intermolecular exon ligation. The enhanced activity of C_4A with 22U RNA probably reflects an advantage of the A·U base pair formed between the A of C_4A and U(22), as indicated by the experiment described below.

Reversal of IVS RNA cyclization takes place by attack of an oligopyrimidine at the phosphate following G(414), reforming a linear IVS (Fig. 1A); it has mechanistic similarities to exon ligation (29). The portion of the IGS that guides reverse cyclization [G(25) to A(28)] is displaced 3′ from the portion that guides splicing (6). Circular IVS RNA formed from wild-type, 22U, and 22A RNAs was purified and tested for reverse cyclization with the four C_4N oligonucleotides (data not shown). The same substrate specificity, $C_5 > C_4U > C_4A > C_4G$, was exhibited in the reopening of all three circles. With all oligonucleotide substrates, the reactivities of the three RNAs were in the same order, wild type > 22U > 22A. Although the efficiency of the reverse cyclization reaction changed, the specificity remained unchanged, consistent with the absence of direct interaction of the oligonucleotides with position 22. Therefore, it is likely that in *trans* splicing, where the specificity changes, there is direct interaction of the oligonucleotides with position 22.

DISCUSSION

Two of the most highly conserved nucleotides in group I precursor RNAs are the U at position −1 and the G opposite it in helix P1 (Fig. 1A). On the basis of the conservation and

FIG. 3. Demonstration that increasing the guanosine substrate concentration enhances splicing of mutant RNAs. (A) Reaction of -1C and the corresponding wild-type (WT) pre-rRNA. (B) Reaction of -1A and the corresponding wild-type pre-rRNA. Pre-rRNAs were incubated for 30 min at 30°C under standard splicing conditions with 0.05, 0.5, or 5 mM guanosine (G), inosine (I), or 2AP. Samples were analyzed on a denaturing 4% polyacrylamide gel. Bands: CIR, circular intron; PRE, precursor; LE, ligated exons; IVS, linear intron; L-19, linear intron missing 19 nt from the 5′ end, formed by a cascade of cyclization and site-specific hydrolysis reactions (36); 5′ EXON, free 5′ exon; 5′ EX-IVS, a two-thirds molecule consisting of 5′ exon and intron formed by hydrolysis at the 3′ splice site. Starting material showed only the single band of unreacted precursor (not shown).

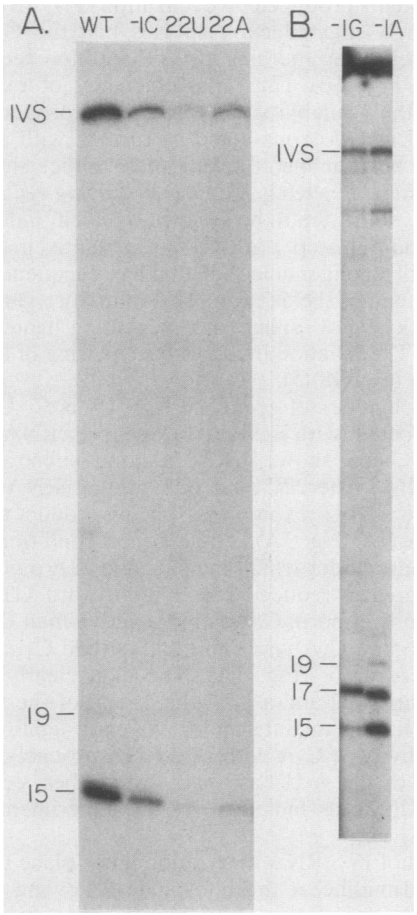


FIG. 4. Inaccuracy in the first step of splicing caused by purine substitution at position −1. (A) Unlabeled -1C, 22U, and 22A pre-rRNAs (0.1 μM) were incubated with 5 μM [α-^{32}P]GTP for 10 min under cyclization conditions [100 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 30 mM Tris, pH 7.5, 42°C]. The products were resolved on a denaturing 20% polyacrylamide gel. (B) Unlabeled -1G and -1A precursor RNAs (0.03 μM) were incubated with 6.25 μM [α-^{32}P] GTP for 60 min under cyclization conditions. Both lanes contain a new, 17-nt product. Additional products seen in the −1 purine lanes may correspond to utilization of cryptic sites in the 3′ exon, as determined previously for deletion mutants of the 5′ exon (24). Their occurrence appears to be condition dependent. WT, Wild-type RNA.

location of these nucleotides, a role in specifying the site of GTP addition has been suggested (17, 29, 30). However, no direct test of the contribution of these nucleotides has been reported.

We now show that any mutation of the conserved U in the *Tetrahymena* group I RNA greatly diminishes GTP addition at the 5′ splice site. The mutations activate cryptic 5′ splice sites that restore a U·G juxtaposition (see below). Mutation of the conserved G in the IGS also diminishes GTP addition. In this case, the interpretation is complicated by the possibility that the adjacent G, G(23), substitutes for the mutated G(22) in directing GTP addition; therefore, we may be underestimating the importance to this reaction of the presence of a G in the IGS. Whether or not G(23) can substitute, the general conclusion is the same: both nucleotides of the U·G contribute to 5′ splice site recognition. Recognition of the U·G by the catalytic core of the intron presumably brings the 5′ splice site in proximity to the GTP-binding site (19).

The same arguments of phylogenetic conservation and location in the structure make it seem just as obvious that the U·G should contribute to the second step of splicing (Fig. 1A). In this work, we inferred the activity of nucleotide combinations other than the normal U·G by sequencing the ligated exons of mutant pre-rRNAs. We also measured exon ligation activity directly by using a *trans*-splicing assay, which allows step 2 of splicing to be observed in the absence of step 1. Surprisingly, we find that combinations of variant 5′ exons and mutated introns designed to give C·G, C·U, C·A, and A·U all have greater activity than U·G. The conserved U·G pair is not required for the second step, nor is it the optimal combination.

Cryptic 5′ splice sites. Mutation of U(−1) to either A or G leads to use of new 5′ splice sites after U(−3) and, to a minor extent, U(−5). Such cryptic splice sites have been observed previously in this and one other group I intron, and in most cases they are explicable in terms of binding of the nucleotides preceding the new site to the IGS (4, 11, 24). When U(−1) was changed to C, GTP addition after U(−3) was not seen. It seems reasonable that pairing of CUCUCC with the IGS (Fig. 6B) would compete successfully with pairing of CUCU, whereas pairing of CUCUCA or CUCUCG to the IGS (Fig. 6C and E) would be destabilized and compete less well with pairing of CUCU (Fig. 6D and F).

In these models, the 5′ exon assumes a new alignment on the IGS, while the IGS remains fixed relative to the catalytic core. (The catalytic core contains the GTP-binding site and functional groups that activate the splice site phosphate [8].) We have tested one of the alternative possibilities, that the mutation causes displacement of the GTP-binding site while the 5′ exon·IGS remains fixed relative to the rest of the catalytic core. Activation of the 5′ splice-site phosphate can be assessed by site-specific hydrolysis in the absence of GTP (17). For the one mutant tested (-1G), the primary site of specific hydrolysis was the phosphate following U(−3) (E. T. Barfod, Ph.D. thesis, University of Colorado, Boulder, 1988). Thus, the mutation is not causing the GTP-binding site to be displaced relative to the rest of the catalytic core.

It should be noted that splicing to the cryptic site is very slow (the rate constant of 0.015 min^{-1} in Table 1 includes splicing to both the normal and the cryptic sites). It is therefore much easier to detect splicing to the cryptic site in a mutant with a low background of normal splicing than in the wild-type RNA. We have not seen splicing to the cryptic site in wild-type RNA, but we have not used methods sensitive enough to detect 0.1% of the GTP addition occurring at this site.

GTP addition after A and G. In the -1A mutant, GTP addition occurs to a phosphate following an A, although at a rate 1,000-fold lower than it occurs to the same site when it follows a U. In the -1G mutant, GTP addition occurs to a phosphate following a G; considering the rate data of Fig. 2 and the product distribution (ratio of 15-mer to 17-mer in Fig. 4B), the rate must be even lower than in -1A. Nevertheless, the ribozyme continues to recognize this position in the P1 helix even in the absence of a U·G pair. Moreover, additional disruption of the helix beyond the splice site, as in the -1G/AA mutant, does not alter this specificity but leads to a small further decrease in efficiency relative to the -1G RNA in all reactions tested (Fig. 2B and data not shown).

Although G(22) on the IGS strand is still present in these mutants, it cannot by itself account for the continuation of

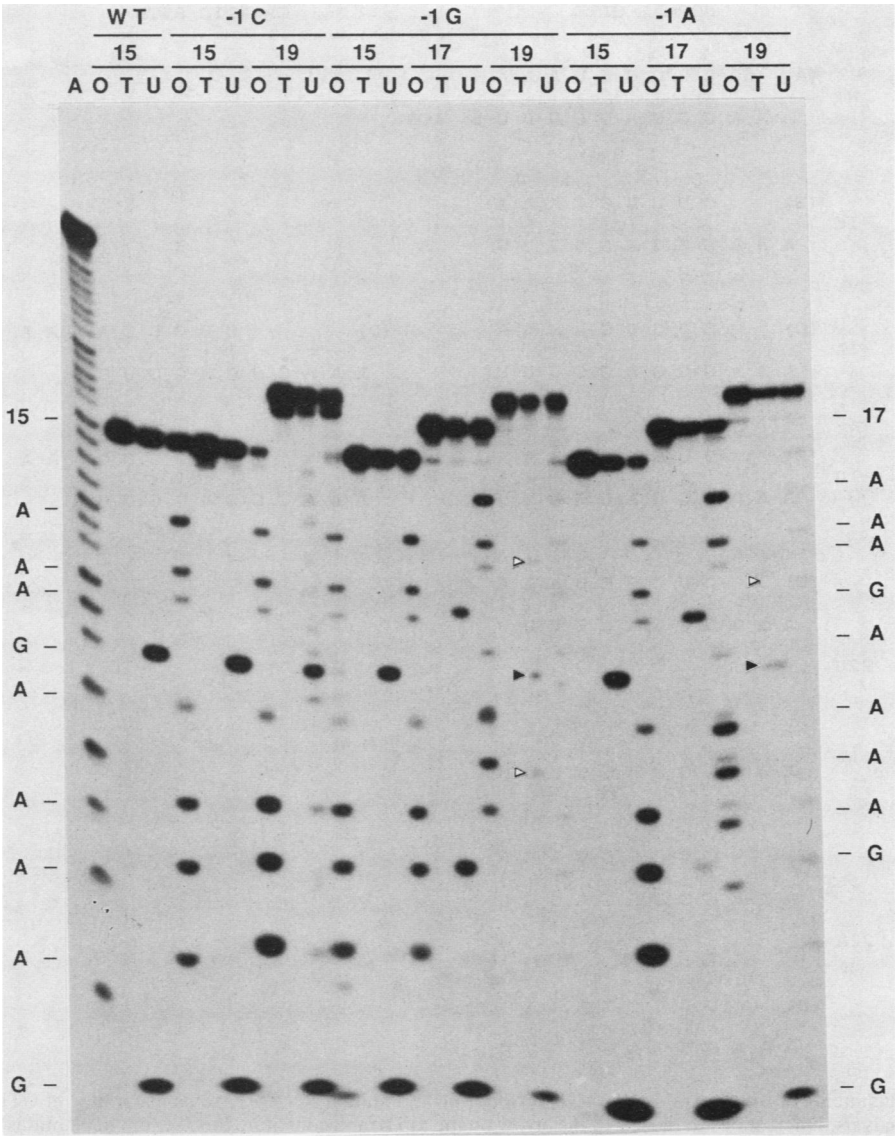


FIG. 5. Sequence analysis of the cyclization products of wild-type (WT), -1C, -1G, and -1A RNAs. The 5'-GTP-labeled oligonucleotide products of splicing and intron cyclization were purified from gels similar to those shown in Fig. 4. The oligonucleotides were then subjected to electrophoresis on a 20% polyacrylamide sequencing gel without additional treatment (lanes O), after partial hydrolysis by RNase T_1, which cleaves after G residues (lanes T), or after partial hydrolysis by RNase U_2, which cleaves after A residues (lanes U). Lane A, Alkaline ladder to provide oligonucleotide size markers. In -1G and -1A RNAs, the 19-nt product had a mixed sequence: bands diagnostic of GTP addition after U(−1) and cyclization after U(19) (▶) and bands diagnostic of GTP addition after U(−5) and cyclization after U(15) (▷).

splicing to position −1. The argument is as follows. There is also a G in the adjacent position in the IGS [G(23)], and it forms a C·G pair, the second-best combination for splicing, yet there is no detectable GTP addition to the phosphate following C(−2).

Except for the U·G at the reaction site, the base pairs in P1 can be substituted without significant effect on the efficiency of GTP addition (5, 34). These other base pairs are therefore unlikely to serve as recognition elements. This leaves the position of a base within the P1 helix as the other determinant of the 5′ splice site. Binding of the phosphodiester backbone of P1 by the catalytic core (5, 19) would be expected to place certain base pairs of P1 close to the GTP-binding site. This could account for the positional dependence. Precise alignment would then be accomplished by specific recognition of the U·G pair.

Some group I introns have more than one U·G pair in P1. The *Tetrahymena* pre-rRNA has U·G for base pairs 2 and 6 of P1, only the latter of which is used. The phage T4 *td* pre-mRNA has U·G for base pairs 3, 7, and 8; the one at position 7 is used (28). These examples emphasize that a minimum of two elements, U·G and position in P1, are necessary to explain 5′ splice site choice.

High K_m for GTP. Mutants at positions −1 and 22 have the highest reported K_m values for GTP. In $(NH_4)_2SO_4$ conditions, the mutants had K_ms 10-fold larger than those of the wild-type RNA (Table 1). In $NH_4C_2H_3O_2$ conditions, the differences were as much as 100-fold (1). A previous K_m mutant showed only a fourfold increase (33).

One way in which the alteration of the U·G pair could produce a large increase in K_m would be if the U·G itself provides a direct binding interaction to the GTP substrate.

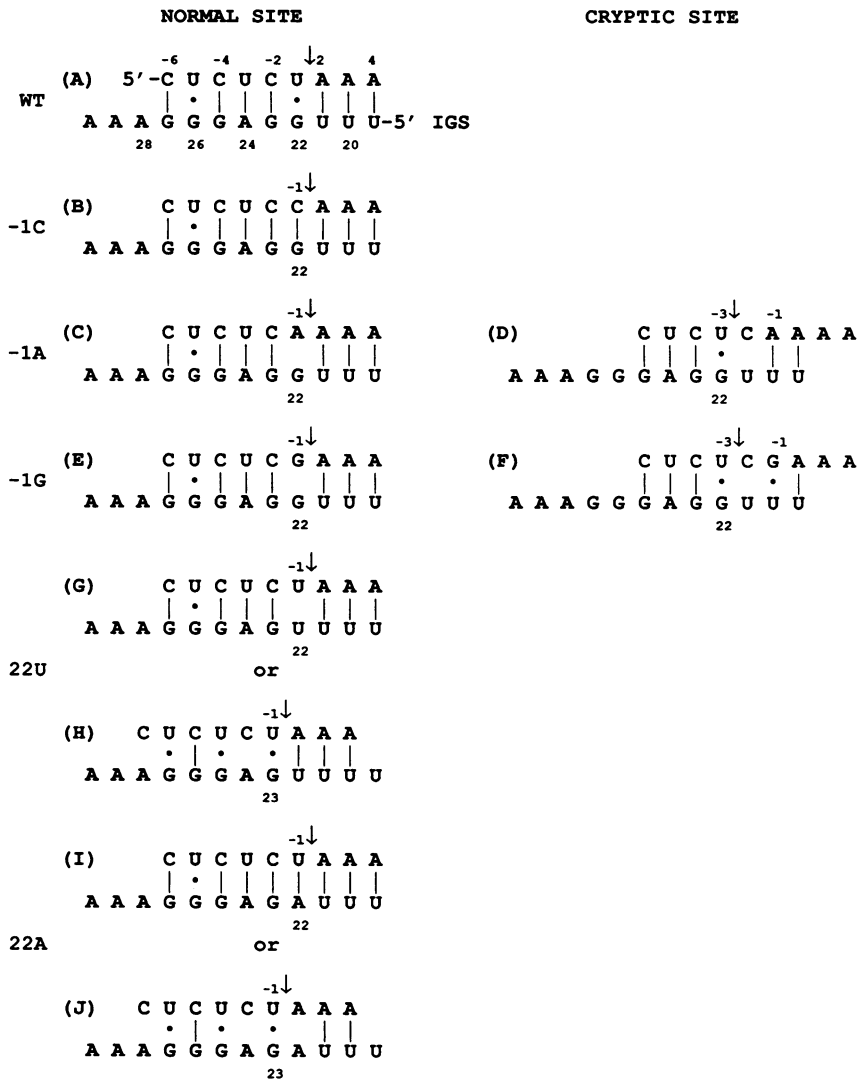


FIG. 6. Proposed alignments of the 5′ exon on the IGS in wild-type and mutant RNAs. Arrows mark sites of GTP addition confirmed by nucleotide sequence analysis. Nucleotides to the left of the arrow on the top strand constitute the 5′ exon; nucleotides to the right of the arrow and those on the lower strand are within the intron.

This was one of the models considered by Inoue et al. (17). To test this model, we tried splicing the −1 mutant RNAs with inosine (which is missing the amino group of guanosine) and 2AP (which is missing the keto oxygen and the imino proton of guanosine). The functional groups that are absent in inosine and 2AP are known to be involved in substrate binding (2). Thus, a mutant precursor no longer able to contact some of these functional groups might, for example, utilize inosine and guanosine equally as substrates. No such specificity change was observed (Fig. 3). We therefore conclude that the mutations are perturbing the binding of GTP indirectly. The scissile phosphate following U(−1) must necessarily be in proximity to the GTP-binding site, so it is not surprising that substitution of other nucleotides would affect GTP binding.

Why U·G? What is it about the conserved U·G that is recognized by the catalytic core? The major possibilities are the conformation or stability of the U·G pair, the particular functional groups provided by these bases, or both (Barfod, Ph.D. thesis). A U·G pair was originally proposed to have the 3-imino proton and 2-oxo group of uracil forming hydrogen bonds with the 6-oxo and 1-imino proton of guanine, altering the relative displacement of the glycosidic linkages from the helix axis by 0.25 nm (12). X-ray crystallography confirmed this base-pairing scheme for a U·G pair in the acceptor stem of yeast $tRNA^{Phe}$ and also showed that this pair perturbs the RNA helix; the U is positioned closer to the helix axis, and the G is positioned farther away from the axis (20, 25; cf. 32). The nonbonded 4-oxo group of U is oriented into the major groove, and the 2-amino group of G is oriented into the minor groove. Therefore, functional groups are available for noncovalent interaction in a distinctive geometry. Recent work has shown that similarly situated U·G pairs in some tRNAs are recognition points for aminoacyl-tRNA synthetases (15, 21).

In our studies, the base pair that could best substitute for U·G was C·G. Since a C·G should not distort the helix, this result would appear to provide an argument against the helix distortion hypothesis. The lower activity of C·G would then be explained by the C substituting an amino group for the

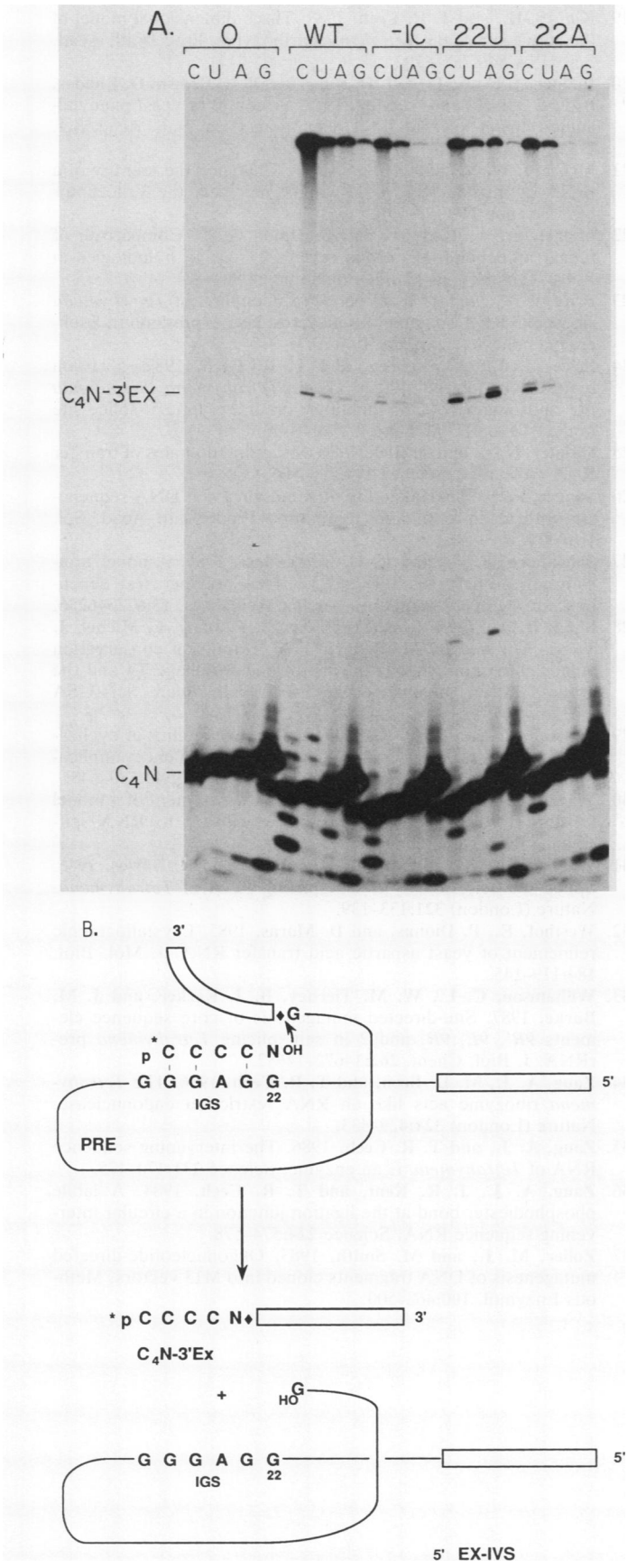


FIG. 7. Enhancement by point mutations at position 22 of reactivity in the second step of splicing. (A) *trans* splicing with oligonucleotides substituting for the free 5′ exon (16). Wild-type (WT), -1C, 22U, and 22U pre-rRNAs (0.05 μM) were incubated with 0.025 μM C_4N (N is C, U, A, or G, as indicated above each lane) for 10 min at 42°C in 100 mM $(NH_4)_2SO_4$–10 mM $MgCl_2$–30 mM Tris, pH 7.5. In lanes O, no pre-rRNA was added. Pre-rRNAs were transcribed from *Hin*dIII-cut plasmid to give a short 3′ exon of 29 nt and contained no radiolabel. C_4N pentanucleotides were 5′ end labeled with polynucleotide kinase and [γ-^{32}P]ATP. Products were analyzed on a denaturing 20% polyacrylamide gel, an autoradiograph of which is shown. Products labeled C_4N-3′EX contain the oligonucleotide C_4N covalently joined to the 5′ end of the 3′ exon. Radiolabeled bands near the top of the gel are products of reverse cyclization of the IVS (see text), the circular IVS being formed from the precursor during incubation (17). Radiolabeled bands near the bottom of the gel are the input C_4Ns, some products resulting from the enzymatic activity of the intron (35), and some unincorporated [γ-^{32}P]ATP. (B) Mechanism of *trans* splicing is an intermolecular version of the second step of splicing (Fig. 1A). Oligonucleotides that can base pair to the IGS can displace the natural 5′ exon and directly attack the 3′ splice site. Only one of the possible alignments of the oligonucleotide C_4N on the IGS is illustrated.

4-oxo group of U or by C·G being too stable for optimum reactivity. However, either the imino tautomer of C or a C protonated at N-3 could in principle pair with G to give exactly the same structure as a U·G wobble pair. It is therefore premature to make a firm decision between the functional group and helix distortion hypotheses.

What determines step 2? On the basis of the exon ligation and *trans*-splicing results, it appears that formation of a standard base pair between the last nucleotide of the 5′ exon and a G in the IGS is not required for the second step of splicing. 5′ exons terminating in A, G, C, or U are competent for exon ligation; the unreactivity of 5′ exons ending in G in *trans* splicing is not an exception that can be given much weight, because oligonucleotides ending in G bind in a special manner (7). The ability of exon ligation to proceed with any nucleotide at the 3′ end of the 5′ exon was first proposed by Waring et al. (31) on the basis of their studies of an aberrant intramolecular ligation reaction of the *Tetrahymena* pre-rRNA that produced a circular 3′ exon species. Our conclusions concerning normal exon ligation are consistent with those of Waring et al. (31).

In our initial description of the *trans*-splicing reaction, we reported that the dinucleotides CU and CC are active but that CA is not (16). The dinucleotides do not have multiple C residues to anchor them to the IGS. Unlike the longer 5′ exons, these dinucleotides might need to pair to the IGS at both positions.

If recognition of a U·G pair is not a major factor in step 2 of splicing, what are the determinants of the reaction? We favor the idea that 5′ exon·IGS base pairing holds the 3′ end of the 5′ exon in the vicinity of the 3′ splice site (31) and that activation of the 3′ splice site phosphate (17, 36) contributes to catalysis and to specificity.

ACKNOWLEDGMENTS

We thank Alice Sirimarco for help in preparing the manuscript and Dianne Lorenz for drafting some of the figures.

This work was supported by Public Health Service grant GM28039 from the National Institutes of Health. T.R.C. is an American Cancer Society Professor.

LITERATURE CITED

1. **Barfod, E. T., and T. R. Cech.** 1988. Deletion of nonconserved helices near the 3′ end of the rRNA intron of *Tetrahymena thermophila* alters self-splicing but not core catalytic activity. Genes Dev. **2**:652–663.
2. **Bass, B. L., and T. R. Cech.** 1984. Specific interaction between the self-splicing RNA of *Tetrahymena* and its guanosine substrate: implications for biological catalysis by RNA. Nature (London) **308**:820–826.
3. **Been, M. D., E. T. Barfod, J. M. Burke, J. V. Price, N. K. Tanner, A. J. Zaug, and T. R. Cech.** 1988. Structures involved in *Tetrahymena* rRNA self-splicing and RNA enzyme activity. Cold Spring Harbor Symp. Quant. Biol. **52**:147–157.
4. **Been, M. D., and T. R. Cech.** 1985. Sites of circularization of the *Tetrahymena* rRNA IVS are determined by sequence and influenced by position and secondary structure. Nucleic Acids Res. **13**:8389–8408.
5. **Been, M. D., and T. R. Cech.** 1986. One binding site determines sequence specificity of *Tetrahymena* pre-rRNA self-splicing, trans-splicing and RNA enzyme activity. Cell **47**:207–216.
6. **Been, M. D., and T. R. Cech.** 1987. Selection of circularization sites in a group I IVS RNA requires multiple alignments of an internal template-like sequence. Cell **50**:951–961.
7. **Been, M. D., and T. R. Cech.** 1988. RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme. Science **239**:1412–1416.
8. **Cech, T. R.** 1987. The chemistry of self-splicing RNA and RNA enzymes. Science **236**:1532–1539.
9. **Cech, T. R.** 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene **73**:259–271.
10. **Cech, T. R., A. J. Zaug, and P. J. Grabowski.** 1981. In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell **27**:487–496.
11. **Chandry, P. S., and M. Belfort.** 1987. Activation of a cryptic 5′ splice site in the upstream exon of the phage T4 *td* transcript: exon context, missplicing, and mRNA deletion in a fidelity mutant. Genes Dev. **1**:1028–1037.
12. **Crick, F. H. C.** 1966. Codon-anti-codon pairing: the wobble hypothesis. J. Mol. Biol. **19**:548–555.
13. **Davies, R. W., R. B. Waring, J. A. Ray, T. A. Brown, and C. Scazzocchio.** 1982. Making ends meet: a model for RNA splicing in fungal mitochondria. Nature (London) **300**:719–724.
14. **Donis-Keller, H., A. M. Maxam, and W. Gilbert.** 1977. Mapping adenines, guanines, and pyrimidines in RNA. Nucleic Acids Res. **4**:2527–2538.
15. **Hou, Y.-M., and P. Schimmel.** 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. Nature (London) **333**:140–145.
16. **Inoue, T., F. X. Sullivan, and T. R. Cech.** 1985. Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: oligonucleotides can function as 5′ exons. Cell **43**:431–437.
17. **Inoue, T., F. X. Sullivan, and T. R. Cech.** 1986. New reactions of the ribosomal RNA precursor of *Tetrahymena* and the mechanism of self-splicing. J. Mol. Biol. **189**:143–165.
18. **Inouye, S., and M. Inouye.** 1987. Oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA, p. 181–204. *In* S. Narang (ed.), DNA and RNA synthesis. Academic Press, Inc., New York.
19. **Kim, S.-H., and T. R. Cech.** 1987. Three-dimensional model of the active site of the *Tetrahymena* ribozyme. Proc. Natl. Acad. Sci. USA **84**:8788–8792.
20. **Ladner, J. E., A. Jack, J. D. Robertus, R. S. Brown, D. Rhodes, B. F. C. Clark, and A. Klug.** 1975. Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. Proc. Natl. Acad. Sci. USA **72**:4414–4418.
21. **McClain, W. H., and K. Foss.** 1988. Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end. Science **240**:793–796.
22. **Michel, F., A. Jacquier, and B. Dujon.** 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie **64**:867–881.
23. **Price, J. V., and T. R. Cech.** 1985. Coupling of *Tetrahymena* ribosomal RNA splicing to β-galactosidase expression in Escherichia coli. Science **228**:719–722.
24. **Price, J. V., J. Engberg, and T. R. Cech.** 1987. 5′ Exon requirement for self-splicing of the *Tetrahymena thermophila* pre-ribosomal RNA and identification of a cryptic 5′ splice site in the 3′ exon. J. Mol. Biol. **196**:49–60.
25. **Quigley, G. J., and A. Rich.** 1976. Structural domains of transfer RNA molecules. Science **194**:796–806.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Schenborn, E. T., and R. C. Mierendorf.** 1985. A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. Nucleic Acids Res. **13**:6223–6236.
28. **Shub, D. A., J. M. Gott, M.-Q. Xu, B. F. Lang, F. Michel, J. Tomaschewski, and M. Belfort.** 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc. Natl. Acad. Sci. USA **85**:1151–1155.
29. **Sullivan, F. X., and T. R. Cech.** 1985. Reversibility of cyclization of the *Tetrahymena* rRNA intervening sequence: implication for the mechanism of splice-site choice. Cell **42**:639–648.
30. **Waring, R. B., and R. W. Davies.** 1984. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing—a review. Gene **28**:277–291.
31. **Waring, R. B., P. Towner, S. J. Minter, and R. Davies.** 1986. Splice-site selection by a self-splicing RNA of *Tetrahymena*. Nature (London) **321**:133–139.
32. **Westhof, E., P. Dumas, and D. Moras.** 1985. Crystallographic refinement of yeast aspartic acid transfer RNA. J. Mol. Biol. **184**:119–145.
33. **Williamson, C. L., W. M. Tierney, B. J. Kerker, and J. M. Burke.** 1987. Site-directed mutagenesis of core sequence elements *9R′*, *9L*, *9R*, and *2* in self-splicing *Tetrahymena* pre-rRNA. J. Biol. Chem. **262**:14672–14682.
34. **Zaug, A. J., M. D. Been, and T. R. Cech.** 1986. The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. Nature (London) **324**:429–433.
35. **Zaug, A. J., and T. R. Cech.** 1986. The intervening sequence RNA of *Tetrahymena* is an enzyme. Science **231**:470–475.
36. **Zaug, A. J., J. R. Kent, and T. R. Cech.** 1984. A labile phosphodiester bond at the ligation junction in a circular intervening sequence RNA. Science **224**:574–578.
37. **Zoller, M. J., and M. Smith.** 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Methods Enzymol. **100**:468–500.