Marel, G. A., & van Boom, J. H. (1988) *Biochemistry 27*, 58-67.
Sugiyama, H., Kawabata, H., Fujiwara, T., Dannoue, Y., & Saito, I. A. (1990) *J. Am. Chem. Soc. 112*, 5252-5257.
von Sonntag, C. (1987) in *The Chemical Basis of Radiation Biology*, Taylor & Francis, London.
Wu, J. C., Kozarich, J. W., & Stubbe, J. (1983) *J. Biol. Chem. 258*, 4694-4697.

# Alternative Secondary Structures in the 5′ Exon Affect both Forward and Reverse Self-Splicing of the *Tetrahymena* Intervening Sequence RNA†

Sarah A. Woodson‡ and Thomas R. Cech*
*Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309-0215*

*Received September 19, 1990; Revised Manuscript Received November 26, 1990*

ABSTRACT: The natural splice junction of the *Tetrahymena* large ribosomal RNA is flanked by hairpins that are phylogenetically conserved. The stem immediately preceding the splice junction involves nucleotides that also base pair with the internal guide sequence of the intervening sequence during splicing. Thus, precursors which contain wild-type exons can form two alternative helices. We have constructed a series of RNAs where the stem-loop in the 5′ exon is more or less stable than in the wild-type precursor, and tested them in both forward and reverse self-splicing reactions. The presence of a stable hairpin in ligated exon substrates interferes with the ability of the intervening sequence to integrate at the splice junction. Similarly, the presence of the wild-type hairpin in the 5′ exon reduces the rate of splicing 20-fold in short precursors. The data are consistent with a competition between unproductive formation of a hairpin in the 5′ exon and productive pairing of the 5′ exon with the internal guide sequence. The reduction of splicing by a hairpin that is a normal feature of rRNA structure is surprising; we propose that this attenuation is relieved in the natural splicing environment.

Efficient and accurate RNA splicing requires that the correct splice sites are recognized by the splicing machinery and that other sites are discriminated against. In nuclear mRNA splicing, there is evidence that exon sequences can affect splice site choice (Reed & Maniatis, 1986; Somasekhar & Mertz, 1985). Since splicing of mRNAs is carried out by a complex assembly of small nuclear RNAs and protein factors, it is difficult to sort out which interaction is altered by changes in exon sequence. In self-splicing of the *Tetrahymena* intervening sequence (IVS)[1] or intron, excision of the intron and ligation of the exons is catalyzed by the intron RNA itself (Kruger et al., 1982). In this comparatively simple RNA-only system, we can understand some of the effects of exon sequences in self-splicing on a molecular level.

The *Tetrahymena* IVS is a member of a class of introns, called group I, which share a common secondary structure and mechanism of splicing [reviewed in Cech (1990)]. The mechanism of self-splicing is outlined in Figure 1. The complete reaction consists of two phosphodiester transesterifications. In the first step, the 5′ splice site is cleaved by addition of an exogenous GTP to the 5′ end of the IVS. In the second step, the 3′ splice site is cleaved, resulting in ligation of the exons and release of linear IVS RNA. In addition to the properly folded intron, the reaction depends only on magnesium ion and guanosine or GTP concentration. Base-pairing between nucleotides in the 5′ exon and a polypurine sequence in the IVS called the internal guide sequence (IGS) is conserved among group I introns (Davies et al., 1982; Michel et al., 1982). This pairing, designated "P1" (Burke et al., 1987), enables recognition of the 5′ splice site and is required for self-splicing (Been & Cech, 1986; Waring et al., 1986). Another pairing, "P10," which joins an adjacent portion of the IGS with nucleotides in the 3′ exon (Davies et al., 1982), aids in 3′ splice site recognition (Michel et al., 1989; Suh & Waring, 1990), although it is neither necessary nor sufficient for that process (Price & Cech, 1988).

We have previously shown that splicing of the *Tetrahymena* IVS is fully reversible (Woodson & Cech, 1989). The reverse splicing reaction follows the same pathway as the forward reaction outlined in Figure 1, but in the opposite direction. The products of the forward reaction, linear IVS and ligated exons, are now the reactants. Integration of the IVS back into the splice junction of the exons produces a molecule which contains the splice site sequences of the precursor. Reverse splicing is promoted by high magnesium ion and RNA concentration, and the absence of GTP. As in the case of the forward reaction, recognition of the substrate by the IVS RNA depends on base-pairing between nucleotides immediately upstream of the splice junction and the IGS. Natural 3′ exon sequences are not essential for intron integration.

The natural splice junction of the *Tetrahymena* IVS is immediately flanked by two stem-loops that are phylogenetically conserved in the mature rRNA (Clark et al., 1984;

† S.A.W. was supported by a postdoctoral fellowship from the American Cancer Society. T.R.C. is an investigator of the Howard Hughes Medical Institute and American Cancer Society Professor.
* Author to whom correspondence should be addressed.
‡ Present address: Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742.

[1] Abbreviations: CHES, 2-(*N*-cyclohexylamino)ethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; HEPES, *N*-(2-hydroxyethyl)-piperazine-*N*′-2-ethanesulfonic acid; IVS, intervening sequence or intron; IGS, internal guide sequence; LE, ligated exon(s); RNase P, ribonuclease P; rRNA, ribosomal ribonucleic acid; Tris, tris(hydroxymethyl)aminomethane; wt, wild type.

0006-2960/91/0430-2042$02.50/0 

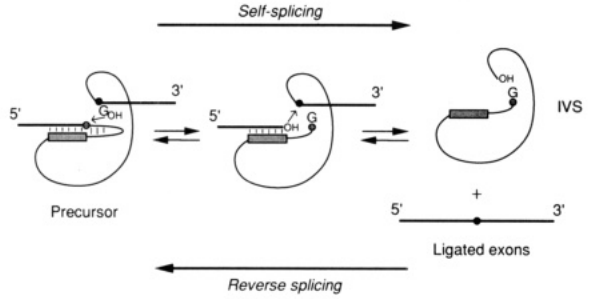


FIGURE 1: Two-step mechanism of splicing in group I introns. In the forward reaction, attack of noncovalently bound GTP at the 5′ splice site (hatched circle) cleaves the 5′ exon. The 3′-hydroxyl of the 5′ exon then attacks at the 3′ splice site (solid circle) in the second step, resulting in excision of a linear IVS and exon ligation. Reverse splicing occurs by the same pathway in the opposite direction (from right to left in the figure), resulting in integration of IVS RNA into the splice junction (solid circle) of the ligated exons. Obligatory base-pairing between nucleotides near the 5′ splice site and a polypurine sequence within the intron called the internal guide sequence (shaded rectangle) is denoted by thin lines.

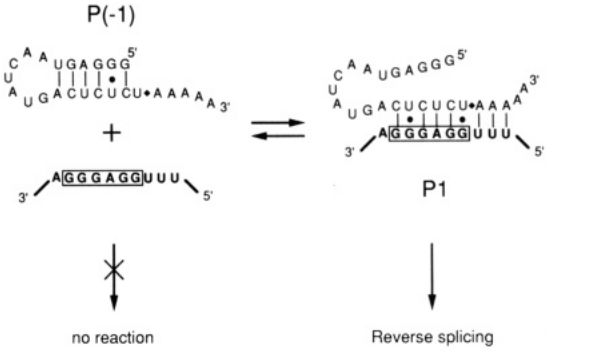


FIGURE 2: Alternative pairing partners for ligated exon substrates in reverse splicing. Ligated exon substrates either can form a hairpin within the 5′ exon as drawn on the left of the figure, and designated P(−1), or can pair with the IVS as shown on the right side (P1). [The P(−1) nomenclature is a departure from the normal convention that paired regions with integral numbers represent conserved secondary structures in group I introns. While the P(−1) structure is a phylogenetically conserved feature of large subunit rRNA, only two other cases are known in which this particular stem–loop precedes the 5′ splice site of a group I intron (see Discussion).] P1 is required for reverse splicing; P(−1) does not lead to products. IVS sequences are shown in boldface type, with the internal guide sequence (IGS) outlined by the box. The splice junction in the ligated exons is denoted by (♦). Watson–Crick base pairs are shown by a line; wobble pairs by (•).

Noller et al., 1981). The 5′ flanking stem involves the same nucleotides of the 5′ exon that must pair with the internal guide sequence of the IVS in forward and reverse splicing. Thus, ligated exon substrates for intron integration which contain the sequences of the stem in the 5′ exon can adopt two alternative conformations as diagrammed in Figure 2. Either they can form the hairpin in the 5′ exon as shown on the left side of the figure, or they can pair with the IGS to form a helix analogous to P1 in the precursor, as shown on the right. For ease of discussion, we will designate the former pairing as P(−1), where −1 denotes a stem in the 5′ exon immediately preceding the 5′ splice site. P(−1) is not expected to contribute to product formation, because the substrate is unable to bind to the IVS in the usual manner.

Previous experiments with ligated exon substrates that contain the sequences of the natural hairpins demonstrated that integration of the IVS is diminished roughly 50-fold over integration into unstructured substrates (Woodson & Cech, 1989). In addition, we were unable to observe integration into longer ligated exons without denaturation of the substrate by glyoxal modification. These results and similar work on the related endonuclease reaction catalyzed by the IVS (Zaug et al., 1986) suggest that the secondary structure in the ligated exons and other target RNAs interferes with recognition by the IVS.

Since these hairpins are a feature of the natural splice junction of the *Tetrahymena* IVS, we wished to examine their effect on the reverse splicing reaction in more detail. 5′ splice site selection by base-pairing with the IGS was better understood than 3′ splice site selection at the time these studies were undertaken. Consequently, we confined our investigation to the 5′ exon stem. Substrate RNAs were synthesized with variations in the 5′ hairpin, and with $A_5$ substituted for the natural 3′ exon. These substrates were then tested in the reverse splicing reaction. The data presented here provide a molecular mechanism to explain our prior observation that certain 5′ exon sequences inhibit integration of the IVS.

We also measured rates of self-splicing in short precursors containing only the 20 nucleotides of the upstream stem–loop. Site-specific changes were made which varied the stability of P(−1). Unexpectedly, the presence of the natural stem in a short precursor lowers the rate of splicing 20-fold, although the wild-type P(−1) has a calculated stability comparable to that of P1. This raises many questions about the relative stabilities of helical segments in a large RNA, and the possibility that conformational switches may influence excision of the IVS from the ribosomal RNA precursor.

## Materials and Methods

*Materials.* Oligodeoxyribonucleotides were synthesized on an Applied Biosystems DNA synthesizer. Unlabeled nucleoside triphosphates were purchased from Pharmacia and radiolabeled nucleotides from New England Nuclear. Restriction enzymes, T4 polynucleotide kinase, and RNA ligase were purchased from New England Biolabs. Reverse transcriptase was purchased from Life Sciences. Phage $T_7$ RNA polymerase was purified by the procedure of Davanloo et al. (1984). Plasmid pTZ18U was from Bio-Rad.

*Reverse Splicing Reactions.* Linear IVS and ligated exon RNAs were prepared by transcription with phage $T_7$ RNA polymerase as described previously (Woodson & Cech, 1989; Zaug et al., 1986), except that synthetic DNA templates for transcription of ligated exon RNA were isolated from 8 M urea–polyacrylamide gels before use in the transcription reaction. Reverse splicing reactions were carried out exactly as described by Woodson and Cech (1989).

*Plasmid Construction and Site-Specific Mutagenesis.* Plasmid pTZIVS+ (a generous gift from Michael Been) contains the *Hin*dIII–*Eco*RI fragment of pβGST7 (Been & Cech, 1986) cloned into pTZ18U (Nolander et al., 1983). Single-stranded deoxyuridine-containing DNA was grown from *Escherichia coli* strain CJ236, and site-specific mutagenesis was carried out by the method of Kunkel (Kunkel, 1985; Kunkel et al., 1987). Plasmid pTZIVSΔ12 was constructed by deleting 12 base pairs from the 5′ exon sequences of pTZIVS+ to give the 20-nucleotide 5′ exon shown for the wild type in Figure 4. A total of 24 base pairs were deleted to give the 8-nucleotide 5′ exon of pTZIVSΔ24, shown for wtΔ24 in Figure 4. In both cases, the *Hin*dIII site of pTZIVS+ was destroyed. Plasmids pTZΔ12-7G, pTZΔ12-14C, pTZΔ12–15A, pTZΔ12-18C, and pTZΔ12-18C:-16C were made by oligonucleotide-directed mutagenesis of pTZIVSΔ12. Plasmid DNA containing the desired sequence changes was identified by dideoxy sequencing using reverse transcriptase, and purified from CsCl gradients.

*Preparation of Precursor RNA.* Precursor RNA was transcribed from the $T_7$ promoter of pTZIVSΔ12 and related plasmids that were linearized with *Eco*RI as described previously (Zaug et al., 1986). Uniformly labeled RNA was transcribed in a 20-μL reaction containing 1 μg of DNA, 2000 units of $T_7$ polymerase, and [α-$^{32}$P]ATP or [α-$^{32}$P]UTP, with incubation at 30 °C for 60 min. Precursor RNA was isolated from 8 M urea–4% polyacrylamide gels, soaked from the gel matrix, and precipitated with ethanol twice to remove contaminating urea and salts.

3′ end-labeled precursors were prepared from 1-mL transcription reactions containing 1 mM each ribonucleoside triphosphate and no radiolabeled triphosphates. Following transcription, the RNA was precipitated with ethanol and passed over a 10-mL Sephadex G-50 column (Pharmacia). Fractions containing RNA were pooled and ethanol-precipitated, the pellet was resuspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA, and the concentration was determined by the absorbance at 260 nm. The RNA was 3′ end-labeled with [5′-$^{32}$P]pCp and RNA ligase (England & Uhlenbeck, 1978). End-labeled precursors were purified by denaturing gel electrophoresis as described above.

*Self-Splicing Reactions.* Uniformly labeled precursor RNA was incubated in 100 mM ammonium sulfate, 50 mM HEPES, pH 7.5, and 5 mM $MgCl_2$ at 50 °C for 2 min to improve renaturation of the IVS, while minimizing loss of starting material through 3′ splice site hydrolysis. The solution was then equilibrated at 30 °C, and reactions were started with the addition of GTP to 100 μM final concentration. Reactions were also carried out in 100 mM ammonium sulfate, 50 mM HEPES, pH 7.5, and 10 mM $MgCl_2$ at 42 °C. Aliquots were removed at various times and added to an equal volume of stop mixture containing 10 M urea and loading dyes. Samples were electrophoresed on denaturing 4% polyacrylamide gels. Radioactivity in each lane was quantitated by using an AMBIS radioanalytic scanner (Automated Microbiology Systems).

The fraction of IVS (excised linear plus circular forms) was determined for each lane according to $f(\text{IVS}) = c(\text{IVS})/[c(\text{pre}) + c(\text{others}) + 1.11c(\text{IVS})]$, where $c$(IVS) is the number of counts in linear and circular IVS bands, $c$(pre) is the number of counts in the precursor band, $c$(others) are the counts in all other reaction products, and the factor 1.11 compensates for the radioactivity in the ligated exons, which were not quantitated. The fraction of spliced products is equal to $f$(IVS) divided by the fraction of the activity calculated to be in the IVS portion of the precursor (0.90 for RNA labeled with ATP or UTP). Observed rates were determined from linear fits to ln [$f$(spliced products)] vs time (minutes), normalized to the fraction of precursor reacted after 4 h of incubation. At long time points, the extent of splicing as judged by $f$(IVS) will be overestimated if a significant amount of 3′ splice site hydrolysis followed by circularization of the IVS has occurred. Data points where a large amount of 5′-exon–IVS and circular IVS had accumulated were not included in the rate determination.

*3′ Splice Site Hydrolysis.* Precursor RNA was incubated at 42 °C in 50 mM CHES, pH 9.0, and 10 mM $MgCl_2$ in the absence of GTP, for up to 225 min. Samples were separated and quantitated as above. Rates were determined from linear fits to ln [$f$(precursor)] vs time for the first 15 min of the time course.

*Competition Experiments.* Splicing reactions were carried out on 3′ end-labeled precursors in a 5-μL volume as described above, but with the addition of 0, 1, 10, or 100 μM unlabeled 5′ exon RNA and 500 μM GTP. The 5′ exon RNA was prepared as described in Woodson and Cech (1989). After 5 min of incubation at 30 °C, 5 μL of stop mixture was added, and samples were electrophoresed on a denaturing 8% polyacrylamide gel.

## Results

*Inhibition of Reverse Splicing with Ligated Exons Containing a Stable Hairpin.* The effect of exon secondary structure on reverse splicing was investigated with a series of substrates that contain 20 nucleotides of natural 5′ exon sequences and $A_5$ substituted for the 3′ exon. These were synthesized by $T_7$ RNA polymerase transcription of synthetic DNA templates (Milligan et al., 1987). In the presence of the IVS, these substrates can potentially adopt two conformations, which are outlined in Figure 2. Either they can form the hairpin within the 5′ exon, P(−1), or 5′ exon nucleotides can base pair with the internal guide sequence to form P1. We have previously shown that pairing of the 5′ exon with the internal guide sequence is required for intron integration (Woodson & Cech, 1989).

A series of sequence changes were made to introduce destabilizing base mismatches in one helix or the other. Alternatively, compensatory changes were made which restored the base-pairing. These substrates are diagrammed schematically in Figure 3A. The double pairings are meant to represent the two alternative pairing possibilities drawn explicitly in Figure 2, and do not represent formation of a triple helix. [This representation is taken from that used to describe alternative pairing in the trp attenuator (Olender et al., 1979).] This is reduced to a plus and minus notation in Figure 3B below, where pairing within the 5′ exon is denoted by P(−1)+ and that between the 5′ exon and the IVS by P1+.

Each $^{32}$P-labeled substrate was tested in the reverse splicing reaction with unlabeled linear IVS under standard conditions. In the wild-type substrate, where both pairings are possible, no product was observed (see Figure 3B). In substrate −18C, a G to C change was made at position −18, outlined by the box, which destabilizes the ligated exon hairpin. This substrate did give the expected reverse splicing product in an amount similar to an unstructured 13-nucleotide substrate (Figure 3B and data not shown). The C to G change at position −4 should destabilize the P1 pairing as well as the P(−1) hairpin, and no products were observed, as expected from previous data on short mismatched substrates (Woodson & Cech, 1989). When base-pairing with the internal guide sequence was restored by a compensatory change at position 25 in the IVS, reverse splicing products were again obtained. In the last reaction, both helices have been restored, so that the pairing opportunities are the same as in the wild-type sequences, but a total of three changes have been made. Once again, no integration products were seen. Thus, substrates which can form an alternative helix within the 5′ exon do not react with the intron RNA, whereas ligated exons which have a destabilized P(−1) hairpin give the expected product.

Since integration of the intron depends on pairing of the 5′ exon with the internal guide sequence, these results indicate that the P(−1) stem in the ligated exons competes effectively with P1, inhibiting the reverse splicing reaction. Because both −18C and −4G:25C combinations give product, the inhibitory effect is due to secondary structure and not to primary sequence changes per se. Eight other substrates with varying stabilities in the ligated exon hairpin were tested in a similar manner (data not shown). Again, product formation was only observed for those with destabilized P(−1) hairpins. Therefore, the ability of a particular RNA to act as a substrate for reverse splicing depends on sequence complementarity between the

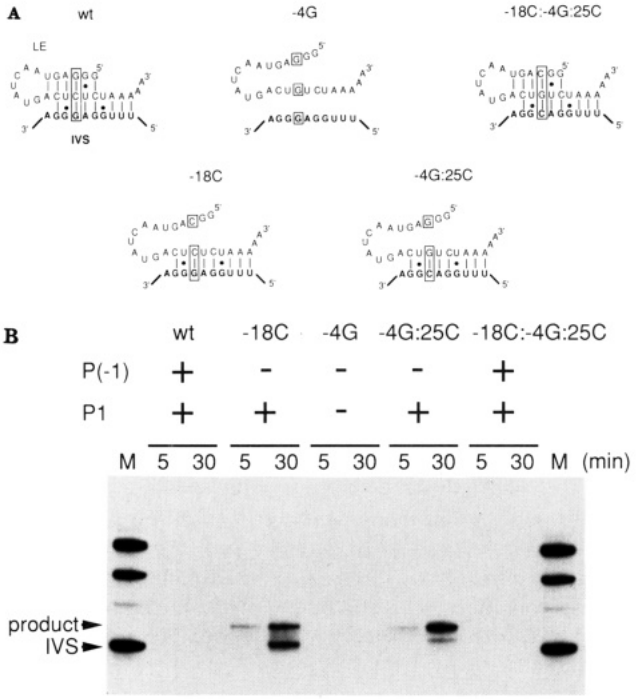


FIGURE 3: Reverse splicing of 25-nt ligated exon substrates. (A) RNA substrates for IVS integration contain a 20-nucleotide 5′ exon and $A_5$ as a 3′ exon. The double pairings indicated by the thin lines are meant to represent the two alternate conformations as shown in Figure 2 and do not represent triple helices. Sequence changes were made at the boxed positions to destabilize the helices as shown. LE denotes ligated exons, shown in lightface type; IVS denotes the intervening sequence RNA, shown in boldface type. Negative numbers refer to exon sequences; positive numbers refer to positions within the IVS. (B) Reverse splicing reactions were carried out on the substrates shown in (A) at 42 °C in 50 mM HEPES, pH 7.5, 100 mM ammonium sulfate, and 25 mM $MgCl_2$. The reaction mixture was electrophoresed on a 4% denaturing polyacrylamide gel. The upper band marked "product" results from the complete reverse splicing reaction. The lower product band contains a shortened form of the IVS inserted into the ligated exons and results from circularization of the intron RNA. Circularization at multiple sites within the intron may cause the length of the lower product to vary (Woodson & Cech, 1989). Time of incubation is noted at the top of the autoradiogram. Lanes M contain uniformly labeled precursor RNA processed to give linear IVS as a marker. (+) and (−) in the P(−1) and P1 rows indicate whether the sequence preceding the integration site can form the hairpin in the ligated exons or pair with the IGS.

5′ exon and the IGS, and availability of the target for pairing with the IGS.

*Forward Self-Splicing of Precursors with an Alternative Stem-Loop in the 5′ Exon.* Since the presence of the natural stem in the ligated exons inhibits the reverse splicing reaction, we wished to examine its effect on the forward splicing reaction. Plasmid pTZIVSΔ12, when linearized with *Eco*RI and transcribed with $T_7$ RNA polymerase, gives rise to a precursor RNA containing precisely the 20 nucleotides that form the 5′ exon hairpin and 29 nucleotides of 3′ exon sequences. A number of variants were then made where the stability of P(−1) was increased or decreased relative to that of the wild type, but sequences that form P1 were left unchanged. The 5′ exons of these precursors are diagrammed in Figure 4. Another plasmid, pTZIVSΔ24, has an eight-nucleotide 5′ exon. As in the case of the reverse splicing reaction, the wild-type precursor can adopt two foldings around the 5′ splice site: either formation of P(−1) as shown on the top left of Figure 4, or formation of P1 as shown on the top right. Again, P1 is required for self-splicing, and the P(−1) folding is expected to be unproductive.

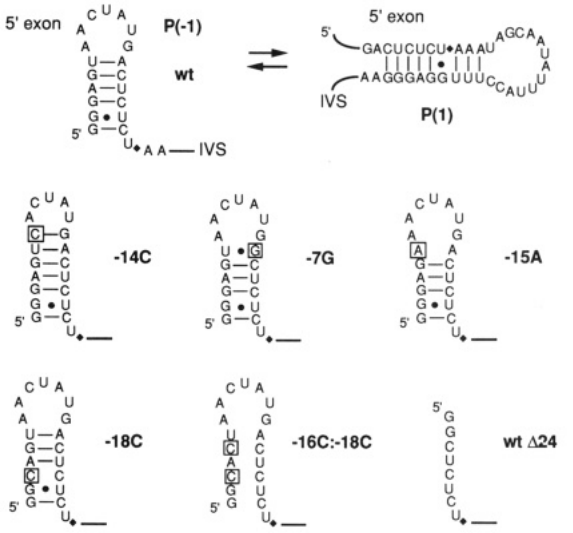


FIGURE 4: 5′ exon stem–loops in precursor variants. 5′ exon sequences of short precursors are as drawn, with sequence changes from the wild type boxed in. (♦) indicates the 5′ splice site; the heavy line represents continuation of the IVS RNA. Alternate helices P(−1) and P1 are shown only for the wild type at the top of the figure. Base pairs drawn as in Figure 2.

Table I: Reactions of Precursor RNA Variants

| RNA | $\Delta G°_{37}$[a] (kcal/mol) | splicing[b] $k_{obs}$ ($min^{-1}$) | hydrolysis[c] $k_{obs}$ ($min^{-1}$) |
|---|---|---|---|
| −14C | −8.0 | 0.004 | 0.21 |
| wt | −6.0 | 0.02 | 0.19 |
| −7G | −5.8 | 0.02 | 0.20 |
| −15A | −3.9 | 0.19 | 0.12 |
| −18C | −2.2 | 0.47 | 0.11 |
| −16C:−18C | +1.6 | 0.37 | 0.10 |
| wtΔ24 | | 0.39 | 0.14 |

[a] Free energy of P(−1) at 37 °C calculated according to Freier et al. (1986). The calculated free energy of P1 is −6.3 kcal/mol. [b] The observed rate of splicing in 100 mM ammonium sulfate, 50 mM HEPES, pH 7.5, and 5 mM $MgCl_2$, 30 °C, with 100 μM GTP. [c] The observed rate of 3′ splice site hydrolysis in 50 mM CHES, pH 9.0, 200 mM NaCl, and 10 mM $MgCl_2$, 42 °C, with no GTP.

Standard free energies of helix formation at 37 °C (Table I) were calculated according to Freier et al. (1986a,b). The precursors are listed in order of descending stability of P(−1). These parameters were determined for duplexes in 1 M NaCl, which roughly approximates helix formation in 10 mM $MgCl_2$ (Williams et al., 1989). Since all of our experiments are carried out in 5–25 mM $MgCl_2$, and the order of splicing efficiency does not change with temperature between 30 and 42 °C, these numbers provide a reasonable way of ranking the stabilities of P(−1) in these precursors. Values at 37 °C are presented because of their lower uncertainty, which is due in part to the few data for entropic and enthalpic contributions of U·G wobble pairs and hairpin loops. It should be emphasized that we only intend these $\Delta G°$'s as a rough guide to facilitate comparison of the hairpins. They should not be mistaken for experimentally determined values.

The rate of self-splicing of each precursor was measured at 30 °C in the presence of near-saturating (100 μM) GTP (Figure 5A and Table I). The wild-type precursor, which is able to form a stable P(−1) stem in the 5′ exon, splices very slowly (0.02 $min^{-1}$), while precursors such as −18C and −16C:−18C, where this hairpin is destablized, react 20-fold faster (0.47 and 0.37 $min^{-1}$). Precursor −15A has a P(−1) stem of intermediate stability and splices at an intermediate rate. Precursor −14C, where P(−1) is predicted to be more stable

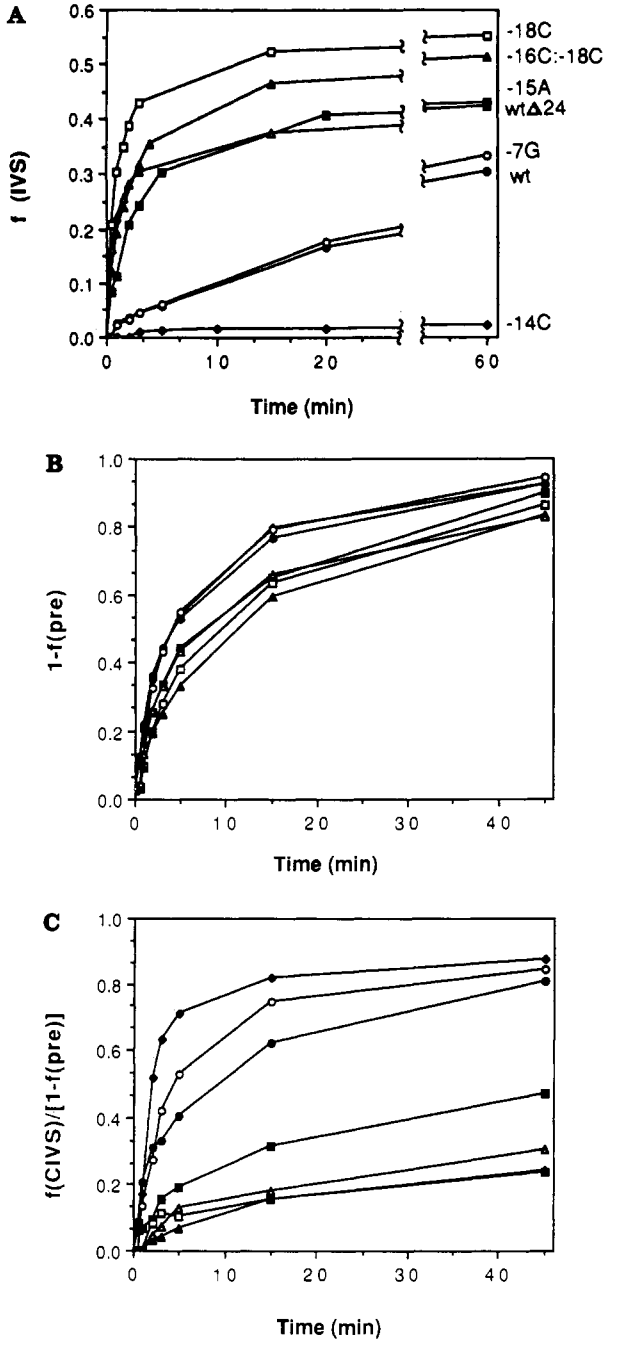


FIGURE 5: (A) Self-splicing. Precursor RNAs containing the 5′ exon sequences shown in Figure 4 were incubated at 30 °C in 50 mM HEPES, pH 7.5, 100 mM ammonium sulfate, 5 mM $MgCl_2$, and 100 μM GTP. Fraction of linear plus circular IVS is plotted vs time. On the basis of the fraction of radioactivity in IVS, 100% splicing would correspond to $f$(IVS) = 0.90. Data points shown here are not adjusted for the total extent of reaction. (B) 3′ splice site hydrolysis. Precursors were incubated at 42 °C in 50 mM CHES, pH 9.0, 200 mM NaCl, and 10 mM $MgCl_2$. 1 − $f$(pre) is plotted versus time. (C) Circularization of the 5′ exon–IVS. Total circular plus linear IVS normalized to the extent of 3′ splice site hydrolysis is plotted vs time for the reactions in part B: (●) wt; (♦) −14C; (○) −7G; (■) −15A; (□) −18C; (▲) −16C:−18C; (△) wtΔ24.

than that of the wild type by an additional G·C pair, barely reacts at all ($k_{obs}$ = 0.004 $min^{-1}$). wtΔ24, which cannot form a hairpin in the 5′ exon, also undergoes rapid splicing. Thus, the rate of reaction in this series of precursors shows negative correlation with the ability to adopt an alternative stem upstream of the 5′ splice site.

Time courses were also carried out in 10 mM $MgCl_2$ at 42 °C, another standard condition which normally increases the rate of splicing over that at 30 °C and promotes circularization of the linear IVS (Cech et al., 1981). Under these conditions, the precursors maintained the same order of reactivity, although all the precursors spliced more rapidly. IVS–3′ exon intermediate did not build up during the course of splicing, even of the less reactive precursors, evidence that attack of GTP at the 5′ splice site during the first step of splicing is rate limiting. Rates of splicing measured at 1 μM GTP gave the same relative rates of splicing of the different precursors.

*3′ Splice Site Hydrolysis of Precursors.* At high pH in the absence of GTP, the 3′ splice site is specifically hydrolyzed, releasing the 3′ exon and leaving a 5′ exon–IVS two-thirds molecule. This reaction is analogous to the second step of splicing, with the 3′ splice site being attacked by hydroxide ion instead of the 3′-hydroxyl of the 5′ exon. Specific hydrolysis of the 3′ splice site depends on the folded structure of the IVS and leaves the 3′-OH, 5′-phosphate termini characteristic of the splicing reaction (Zaug et al., 1984; Inoue et al., 1986).

The precursors above were incubated under standard hydrolysis conditions (10 mM $MgCl_2$, pH 9, 42 °C; Inoue et al., 1986). The extent of reaction vs time is plotted in Figure 5B, and the observed rates listed in Table I. It is immediately apparent that all of the precursors hydrolyze at similar rates, with the slowly splicing precursors such as the wild type consistently reacting 2-fold faster than precursors which splice efficiently, such as −18C. This indicates that intrinsic catalytic properties of the IVS core including the ability to bind guanosine (G414 in the case of 3′ splice site hydrolysis; Inoue et al., 1986; Michel et al., 1989) have not been affected by the changes in the 5′ exon. Furthermore, it supports the idea that misfolding around the 5′ splice site prevents attack of GTP during the first step of splicing but does not affect the ability of the IVS to recognize and activate the phosphodiester bond at the 3′ splice site. Again, IVS–3′ exon RNA, which would result from hydrolysis at the 5′ splice site, is not observed under these conditions, nor are ligated exons, which are a secondary product of 5′ splice site hydrolysis (Inoue et al., 1986).

Following hydrolysis of the 3′ splice site, the 5′ exon–IVS can go on to form circular IVS by attack of the 3′-terminal guanosine, G414, at the 5′ splice site or at circularization sites that precede the internal guide sequence (Inoue et al., 1986). We initially observed an unusual amount of circular IVS appearing very early in the hydrolysis reaction for some of the precursors. Further investigation showed that precursors which contain a stable P(−1) stem circularize rapidly following hydrolysis of the 3′ splice site. The amount of circle normalized to the fraction of hydrolyzed precursor is plotted in Figure 5C. Closer examination of the products indicates that wt, −7G, and −14C precursors give rise almost exclusively to C-15 or C-19 forms of the IVS instead of the full-length circle. These smaller circles result from attack of G414 at the circularization sites, instead of at the 5′ splice site (Zaug et al., 1984). The simplest explanation for these observations is that the stem in the 5′ exon frees the guide sequence for binding of the circularization sites, thereby permitting efficient circularization of the IVS following hydrolysis at the 3′ splice site.

*Competition of Trans and Cis Splicing.* Another way of comparing the ability of these precursors to form P1 is to ask whether free 5′ exon added to the reaction can bind to the internal guide sequence and undergo intermolecular exon ligation or trans-splicing. Folding of the precursor RNA about the 5′ splice site can be treated as an equilibrium between two conformations, I and III, as diagrammed in Figure 6A. For precursors that form a stable hairpin in the 5′ exon, such as

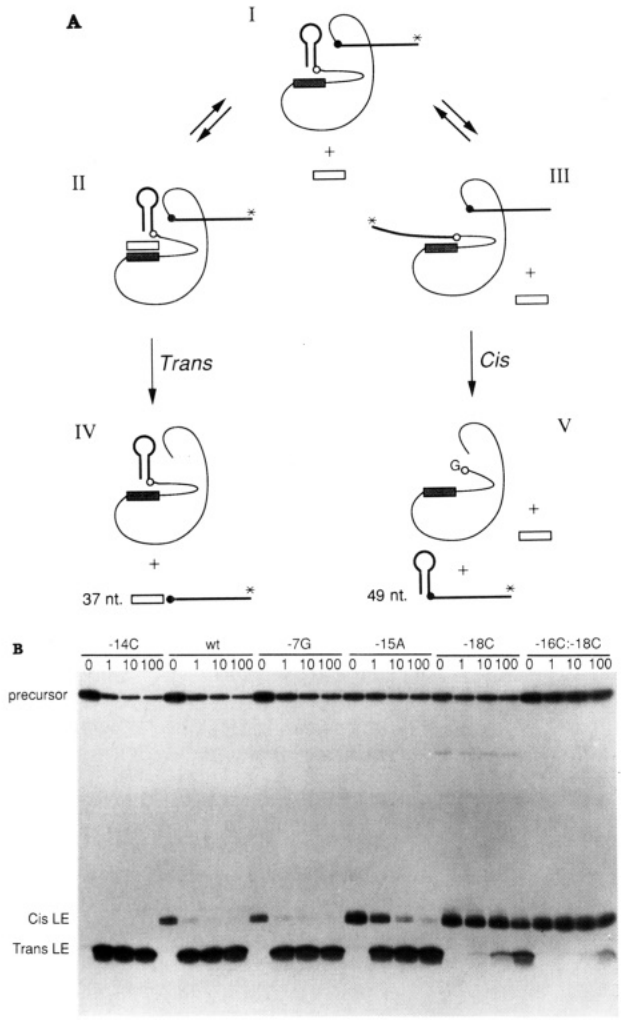


FIGURE 6: Competition between trans and cis splicing. (A) Equilibrium between two folded forms of the precursor (I and III) is shown at the top of the figure. Heavy lines denote exon sequences; thin lines denote IVS sequences. The 5′ splice site is represented by an open circle and the 3′ splice site by a solid circle. An asterisk denotes the 3′ end-label. Pairing of the intramolecular 5′ exon (heavy line) with the internal guide sequence (dark rectangle) in III leads to the cis-splicing products shown in V. Precursors in conformer I can bind free 5′ exon (light rectangle) as shown in II to produce the trans-splicing products shown in IV. GTP is assumed to be noncovalently bound to all the structures shown, except in V where it is attached to the 5′ end of the IVS. (B) Autoradiogram of an 8% polyacrylamide gel showing products of trans- and cis-splicing reactions. −14C, wt, etc. denote the precursor in each reaction. 0, 1, 10, and 100 are the concentrations of added exogenous 5′ exon in micromolar. 3′ end-labeled precursor RNA is indicated at the top of the autoradiogram. Cis LE is the 49-nt product of intramolecular splicing; Trans LE is the 37-nt product of the intermolecular reaction. Cis LE appears as a doublet due to 3′-end heterogeneity of the original $T_7$ transcript.

wt or −14C, this equilibrium should lie toward I. For precursors where this P(−1) hairpin is destabilized relative to P1, such as −18C or −16C:−18C, this equilibrium lies toward III, which can go on to form ligated exons in the usual cis-splicing reaction in the presence of GTP. I is not able to splice, but the internal guide sequence (represented by the dark rectangle) should be available to bind free 5′ exon added to the reaction (light rectangle), as drawn in II. It has been shown that free 5′ exon terminating in a 3′-hydroxyl can undergo trans-splicing by attack at the phosphodiester bond of the 3′ splice site in the usual manner (Inoue et al., 1985). Addition of increasing amounts of competitor 5′ exon to the reaction could be expected to shift the equilibrium in favor of II, increasing the amount of trans-splicing products (IV) over cis products (V).

3′ end-labeled precursors were incubated in the presence of 500 μM GTP under the usual splicing conditions with the addition of 0, 1, 10, or 100 μM free 5′ exon, an octanucleotide of sequence $^{5'}GGCUCUCU_{OH}$. After 5 min of incubation, the reaction mixtures were electrophoresed on a 10% polyacrylamide gel, as shown in Figure 6B. The size of the expected trans-splicing product is 37 nucleotides; the cis-splicing product is 49 nucleotides in length. The precursors are loaded in decreasing order of P(−1) stability from left to right across the gel. −14C, which has a very stable 5′ exon hairpin and whose self-splicing is barely detectable, gave rise to a significant amount of trans product at 1 μM added 5′ exon, but no cis product. As the stability of P(−1) decreases, the relative amount of cis product increases, even at higher concentrations of competitor. In the case of precursors −18C and −16C:−18C, which are normally reactive, very little trans product is formed, and even 100 μM added 5′ exon does not compete out the intramolecular reaction.

The results are then entirely consistent with the model shown in Figure 6A and support the idea that the alternative folding of the 5′ exon prevents it from pairing with the internal guide sequence such as to position the 5′ splice site correctly for G addition. This alternate conformation does not affect the catalytic core of the IVS nor does it interfere with the ability of the intron to recognize the 3′ splice site and carry out the second step of splicing.

## Discussion

The self-splicing activity of group I introns resides within the intron itself (Kruger et al., 1982; Zaug & Cech, 1986). Thus, the major focus of earlier studies has been to understand features of the intron RNA which result in its unique catalytic properties. Previous experiments which looked for exon sequence requirements identified the last six nucleotides of the 5′ exon which pair with the IGS to form P1 (Been & Cech, 1986; Waring et al., 1986; Price et al., 1987; Barfod & Cech, 1989) and the first three nucleotides of the 3′ exon, which can interact with the IGS in the P10 pairing (Davies et al., 1982; Michel et al., 1989). Our data demonstrate that secondary structure modifications in the 5′ exon can cause large changes in the rate of both the forward and reverse self-splicing reactions. The simplest interpretation of the results is that an alternative pairing within the 5′ exon [P(−1)] competes with base-pairing of nucleotides prior to the 5′ splice site with the internal guide sequence (P1). This results in lower rates of reaction, because the catalytic portion of the IVS is prevented from recognizing the phosphodiester bond at the 5′ splice site. In the case of reverse splicing, competition by P(−1) is such that the product is undetectable. Point mutations which destabilize P(−1) increase the rate of both forward and reverse splicing, whereas mutations which stabilize P(−1) decrease the rate of reaction.

Hydrolysis of the 3′ splice site is relatively unaffected by changes in the 5′ exon, indicating that the mutations are not disrupting the folding of catalytically active IVS molecules, and confirming previous observations that 5′ and 3′ splice site recognition is independent (Inoue et al., 1985; Szostak, 1986; Zaug et al., 1986). The trans-splicing reactions provide even better evidence that mutations which alter P(−1) have only a local effect. The extent to which exogenous 5′ exon added to the reaction mixture can compete with the intramolecular 5′ exon for binding to the IGS depends on the stability of P(−1). If P(−1) is strong, the IGS is available for base-pairing, but the intramolecular 5′ exon does not bind to it. Once a 5′ exon is paired to the internal guide, the intron is competent to splice.

In addition to the reverse splicing substrates presented in Figure 3, we also examined 8 other 25-nucleotide substrates which include 5′ exon hairpins identical with those in the precursor series shown in Figure 4. This allows a precise comparison between the forward and reverse reactions. In both cases, there is a relatively sharp transition from substrates with low activity to those with nearly full activity, with the midpoint falling near the stability of P(−1) in precursor −15A, approximately −4 kcal/mol at 37 °C. The amount of integration product formed after 30 min in substrates −18C and −16C:−18C was similar to that of 13-nucleotide, unstructured substrates. No product was observed for the wild type and −7G. A very small amount of product was formed with −15A. This is the same pattern observed for precursor splicing, where −18C, −16C:−18C, and wtΔ24 all react quickly, −15A is intermediate, and wt, −7G, and −14C splice slowly. The estimated stabilities of P(−1) according to Freier et al. (1986b) correlate well with the observed activity of each precursor at 30 °C in 5 mM $MgCl_2$ (see Table I).

*Relative Stabilities of P1 and P(−1).* The transition from reactive to unreactive precursors is clearly modulated by the relative stabilities of P1 and P(−1). The calculated free energies of the two hairpins at 37 °C according to Freier et al. (1986b) are very similar to each other. The $\Delta G°$ of P1 is −6.3 kcal/mol, and $\Delta G°$ of P(−1) is −6.0 kcal/mol. Unexpectedly, the results of the splicing experiments imply that P(−1) is significantly more stable than P1. This apparent discrepancy might be due to inaccuracy in the calculation. For example, the calculation treats all loop sequences as equivalent, but it is known that certain combinations of nucleotides, such as $^5$′CUUCGG, are more stable (Tuerk et al., 1988). A change of 1 kcal/mol in the estimate of $\Delta G°$ for either hairpin would be enough to significantly alter the equilibrium population of the two states; 1 kcal/mol is within the uncertainty of these calculations. Alternatively, P1 and P(−1) could interact differentially with the IVS.

There is evidence for significant stabilization of P1 by interaction with the intron. The binding constants of small substrates to the L-21 form of the IVS have been measured by native gel electrophoresis and by kinetic measurements on the riboendonuclease reaction, which is analogous to the first step of splicing (Pyle et al., 1990; Herschlag & Cech, 1990). The L-21 IVS lacks the 21 nucleotides 5′ of the guide sequence, and P1 is formed intermolecularly by binding small substrate RNAs which contain the sequences of the 5′ exon. From these studies, it was concluded that tertiary interactions with the core of the IVS RNA contribute significant binding free energy beyond that of simple helix formation (an extra 4 kcal/mol at 30 °C; Pyle et al., 1990), in agreement with earlier conclusions of Sugimoto et al. (1988). If these extra interactions are taken into account, pairing with the internal guide sequence would be expected to be significantly more stable than a hairpin in the 5′ exon. Our results, however, clearly show that wild-type P(−1) is preferred over wild-type P1 in short precursors. One possibility is that the IVS core stabilizes both hairpins, although binding of P(−1) is unproductive. P(−1) contains a U·G base pair with the U occupying the fifth position from one of the 5′ ends of the paired regions; this might help it to bind in a mode similar to that of P1 (Doudna et al., 1989).

*Are P1 and P(−1) in Rapid Equilibrium?* Splicing of RNA containing a competing P(−1) stem could be explained by either of two models, which our data do not allow us to distinguish. P1 and P(−1) could interconvert rapidly, in which case the observed rate of splicing would reflect the equilibrium fraction of RNA folded into the active P1 structure. Alternatively, the refolding of the RNA from P(−1) to P1 could be slow enough to be partially or completely rate-limiting for those molecules that began in the P(−1) state.

Unfortunately, few data are available on the rates of hairpin formation, and they have not been experimentally measured for these sequences. Previous estimates of opening and closing rates for RNA hairpins are on the order of $10^4$ $s^{-1}$ at $T_m$ (Gralla et al., 1974; Porschke & Eigen, 1971; Craig et al., 1971). If these rates are corrected to 30 °C with enthalpies calculated from Turner and co-workers for the P(−1) helices, the opening rates would be expected to be between 6 and 60 $min^{-1}$ and 3200 $min^{-1}$ for −18C and the wild-type hairpins, respectively. These rates are much faster than splicing, even allowing for an order of magnitude uncertainty. This would lead one to expect the rapid equilibrium situation. On the other hand, interactions with the IVS core might be expected to slow opening of the 5′ helices. The off rate of $^5$′CUCUCU bound to the IGS is on the order of 1 $min^{-1}$ at 50 °C (Herschlag & Cech, 1990), similar to the rate of splicing. If P(−1) is also stabilized by the intron, it is possible that opening of P(−1) may be a slow step in splicing. Under some conditions, a small burst of product can be observed at the beginning of the time course in self-splicing reactions. The burst accounts for 2–6% of the total RNA in wild-type and −14C precursors, and may represent the fraction of the population where P1 is correctly folded.

*Biological Implications for Pre-rRNA Splicing.* The fact that a hairpin which is naturally present in the 5′ exon of the *Tetrahymena* pre-rRNA has such a deleterious effect on the rate of self-splicing was surprising and puzzling at the outset. These hairpins are found in a conserved region of the ribosomal RNA, and phylogenetic evidence for the pairing is strong (Clark et al., 1984; Noller et al., 1981; Veldman et al., 1981). We originally proposed that the formation of secondary structure in the ligated exon product might serve to prevent reintegration of the IVS, and help drive splicing in the forward direction (Woodson & Cech, 1989). We now find that the secondary structure is already present in the unspliced TZIVSΔ12 RNA and therefore slows excision of the IVS.

These results might appear to contradict earlier data on full-length pre-rRNA synthesized in isolated *Tetrahymena* nuclei, which self-splices at a rate of 0.5 $min^{-1}$ in vitro under similar conditions (Bass & Cech, 1984). More recent work, however, on precursor RNAs which contain 5′ and 3′ exons of *Tetrahymena* rRNA sequence longer than TZIVSΔ12 shows that they splice efficiently, although they still contain the wild-type sequences that potentially form P(−1). Long-range interactions with the rRNA apparently alter the folding around the 5′ splice site and favor the formation of P1, permitting efficient splicing of the IVS. Experiments to identify the precise nature of this interaction are in progress. Self-splicing may consequently be a sensitive probe of RNA structure and conformational changes not only within the IVS itself but also in the domain of rRNA in which it is located.

*Other Group I Introns.* It seems likely that effects of exon structure on splicing will be common for group I introns that inhabit rRNA precursors. When each of these introns is considered in the context of the secondary structure of its rRNA exons, it is apparent that most of them are flanked by phylogenetically proven base-paired regions (Table II). In some cases, the rRNA contains a structure equivalent to P(−1) of the *Tetrahymena* intron, which could compete with the formation of P1, or a hairpin in the 3′ exon which could compete with the formation of P10 and thereby interfere with

Table II: Group I Introns Flanked by Stable rRNA Hairpins

| rRNA | organelle | organism | insertion site[a] | rRNA structure competes with[b] P1 | P10 |
|---|---|---|---|---|---|
| LSU | nuclear | *T. thermophila* | 1925 | + | + |
| | | *P. polycephalum*[c] | | | |
| | | I3 | 1925 | + | na |
| | | I2 | 1949 | +* | na |
| | | I1 | 2449 | + | na |
| | mitochondrial | *S. cerevisiae*[d] | 2449 | + | – |
| | chloroplast | *C. reinhardtii*[e] | 2593 | +* | – |
| | | *C. eugametos*[e] | | | |
| | | I1 | 730 | + | +* |
| | | I2 | 958 | +* | na |
| | | I3 | 1065 | +* | – |
| | | I4 | 1766 | +* | na |
| | | I5 | 1923 | + | + |
| | | I6 | 2596 | +* | +* |
| SSU | nuclear | *A. stipitatus*[f] | 1045 | + | + |
| | | *P. carinii*[g] | 1506 | – | + |
| | chloroplast | *C. moewusii*[e] | 531 | – | na |

[a] Intron follows this nucleotide; numbering systems are those of the *E. coli* 23S and 16S rRNAs (Noller, 1984). [b] +/–, presence/absence of hairpin in 5′ exon that would compete with formation of P1, or hairpin in 3′ exon that would compete with formation of P10; asterisk, the competing hairpin in the rRNA involves 5′ exon sequences pairing to 3′ exon sequences, so it could form only after the entire intron had been transcribed; na, not applicable because the intron lacks convincing P10 pairing. [c] *P.* = *Physarum* (Nomiyama et al., 1981; I3 from J. Varnum and V. Vogt, personal communication). [d] *S.* = *Saccharomyces* (Dujon, 1980; Gutell & Fox, 1988). Many other fungi have similar rRNA introns inserted at the identical position. [e] *C.* = *Chlamydomonas* [*reinhardtii* from Rochaix et al. (1985); *eugametos* from M. Turmel, J. Boulanger, M. N. Schnare, M. W. Gray, and C. Lemieux (submitted for publication); *moewusii* from Durocher et al. (1989)]. [f] *A.* = *Ankistrodesmus* (V. Huss and M. Sogin, personal communication). [g] *P.* = *Pneumocystis* (Sogin & Edman, 1989).

3′ splice site recognition. In other cases (denoted by an asterisk in Table II), the intron interrupts the rRNA precursor adjacent to the loop of a hairpin, so the rRNA structure that could potentially compete with the formation of P1 could only form after the entire intron had been transcribed. The high frequency of competing rRNA structures may simply reflect the high density of hairpins in rRNA, so exon sequences adjacent to the splice sites are likely to have some alternative pairing with other rRNA sequences that competes with their pairing to the IGS.

How are the group I introns that interrupt rRNA transcripts able to overcome the consequences of being located in an unfriendly neighborhood? Although one could imagine that ribosomal proteins bound in the vicinity of the splice sites might somehow facilitate exon–IGS base-pairing, it seems more likely that ribosomal proteins would bind to and thereby stabilize the rRNA secondary structure, enhancing its ability to interfere with splicing. Instead, perhaps the inactive RNA structure and the productive rRNA–intron pairing are in equilibrium, with splicing being restricted to the fraction of the molecules in the latter conformation. As described in the previous section for the *Tetrahymena* intron, there may be other exon sequences or structures that facilitate disruption of the nonproductive structure. Formation and dissolution of these particular rRNA secondary structures could even be a normal feature of ribosome dynamics.

ACKNOWLEDGMENTS

We thank Cheryl Grosshans for synthesis of DNA oligonucleotides and technical advice on site-directed mutagenesis. We also thank Arthur Zaug for preparation of $T_7$ polymerase, Michael Been for providing pTZIVS+ DNA, Dan Celander for critical reading of the manuscript, and the W. M. Keck Foundation for generous support of RNA science on the Boulder campus.

REFERENCES

Barfod, E. T., & Cech, T. R. (1989) *Mol. Cell. Biol. 9*, 3657–3666.
Bass, B., & Cech, T. R. (1984) *Nature 308*, 820–826.
Been, M. D., & Cech, T. R. (1986) *Cell 47*, 207–216.
Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W., & Tabak, H. F. (1987) *Nucleic Acids Res. 15*, 7217–7221.
Cech, T. R. (1990) *Annu. Rev. Biochem. 59*, 543–568.
Cech, T. R., Zaug, A. J., & Grabowski, P. J. (1981) *Cell 27*, 487–496.
Clark, C. G., Tague, B. W., Ware, V. C., & Gerbi, S. A. (1984) *Nucleic Acids Res. 12*, 6197–6220.
Craig, M. E., Crothers, D. M., & Doty, P. (1971) *J. Mol. Biol. 62*, 383–401.
Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035–2039.
Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A., & Scazzocchio, C. (1982) *Nature 300*, 719–724.
Doudna, J. A., Cormack, B. P., & Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402–7406.
Dujon, B. (1980) *Cell 20*, 185–197.
Durocher, V., Gauthier, A., Bellemare, G., & Lemieux, C. (1989) *Curr. Genet. 15*, 277–282.
England, T. E., & Uhlenbeck, O. C. (1978) *Nature 275*, 560–561.
Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986a) *Biochemistry 25*, 3209–3213.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986b) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373–9377.
Gralla, J., Steitz, J. A., & Crothers, D. M. (1984) *Nature 248*, 204–208.
Gutell, R. R., & Fox, G. E. (1988) *Nucleic Acids Res. 16*, r211.
Herschlag, D., & Cech, T. R. (1990) *Biochemistry 29*, 10172–10180.
Inoue, T., Sullivan, F. X., & Cech, T. R. (1985) *Cell 43*, 431–437.
Inoue, T., Sullivan, F. X., & Cech, T. R. (1986) *J. Mol. Biol. 189*, 143–165.
Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982) *Cell 31*, 147–157.
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488–492.
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.
Michel, F., Jacquier, A., & Dujon, B. (1982) *Biochimie 64*, 867–881.
Michel, F., Hanna, M., Green, R., Bartel, D. P., & Szostak, J. (1989) *Nature 342*, 391–395.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.
Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., & Woese, C. R. (1981) *Nucleic Acids Res. 9*, 6167–6189.
Nomiyama, H., Kuhara, S., Kukita, T., Otsuka, T., & Sakaki, Y. (1981) *Nucleic Acids Res. 9*, 5507–5520.
Norander, J., Kempe, T., & Messing, J. (1983) *Gene 26*, 101–106.

Olender, D. L., Zurawski, G., & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5524-5528.
Porschke, D., & Eigen, M. (1971) *J. Mol. Biol. 62*, 361-381.
Price, J. V., & Cech, T. R. (1988) *Genes Dev. 2*, 1439-1447.
Price, J. V., Engberg, J., & Cech, T. R. (1987) *J. Mol. Biol. 196*, 49-60.
Pyle, A. M., McSwiggen, J., & Cech, T. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8187-8191.
Reed, R., & Maniatis, T. (1986) *Cell 46*, 681-690.
Rochaix, J.-D., Rahire, M., & Michel, F. (1985) *Nucleic Acids Res. 13*, 975-984.
Sogin, M. L., & Edman, J. C. (1989) *Nucleic Acids Res. 17*, 5349-5359.
Somasekhar, M. B., & Mertz, J. E. (1985) *Nucleic Acids Res. 13*, 5591-5609.
Sugimoto, N., Kierzek, R., & Turner, D. H. (1988) *Biochemistry 27*, 6384-6392.
Suh, E. R., & Waring, R. B. (1990) *Mol. Cell. Biol. 10*, 2960-2965.
Szostak, J. W. (1986) *Nature 322*, 83-86.
Tuerk, C., Gauss, P., Thormer, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Brody, E. N., & Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364-1368.
Veldman, G. M., Klootwijk, J., de Regt, V. C. H. F., Planta, R. H., Branlant, C., Krol, A., & Ebel, J.-P. (1981) *Nucleic Acids Res. 9*, 6935-6952.
Waring, R. B., Towner, P., Minter, S. J., & Davies, R. W. (1986) *Nature 321*, 133-139.
Williams, A. P., Longfellow, C. E., Freier, S. M., Kierzek, R., & Turner, D. J. (1989) *Biochemistry 28*, 4283-4291.
Woodson, S. A., & Cech, T. R. (1989) *Cell 57*, 335-345.
Zaug, A. J., & Cech, T. R. (1986) *Science 231*, 470-475.
Zaug, A. J., Kent, J. R., & Cech, T. R. (1984) *Science 224*, 575-578.
Zaug, A. J., Been, M. D., & Cech, T. R. (1986) *Nature 324*, 429-433.

# Location of the Disulfide Bonds in Human Plasma Prekallikrein: The Presence of Four Novel Apple Domains in the Amino-Terminal Portion of the Molecule†

Brad A. McMullen, Kazuo Fujikawa, and Earl W. Davie*
*Department of Biochemistry, University of Washington, Seattle, Washington 98195*
*Received September 24, 1990; Revised Manuscript Received November 19, 1990*

ABSTRACT: The location of 16 of the 18 disulfide bonds in human plasma prekallikrein was determined by amino acid sequence analysis of cystinyl peptides produced by chemical and enzymatic digestions. A unique structure, named the apple domain, was established for each of the four tandem repeats in the amino-terminal portion of the molecule. The apple domains (90 or 91 amino acids) contain 3 highly conserved disulfide bonds linking the first and sixth, second and fifth, and third and fourth half-cystine residues present in each repeat. The fourth tandem repeat contains an extra disulfide bond that forms a second small loop within the apple domain. The carboxyl-terminal portion of plasma prekallikrein containing the catalytic region of the molecule was found to have disulfide bonds located in positions similar to those of other serine proteases.

Plasma prekallikrein, a precursor of plasma kallikrein, circulates in blood as an equimolar complex with high molecular weight kininogen (Mandle et al., 1976). It participates in the generation of kinin (Werle, 1955; Margolis, 1958; Laake & Vennerod, 1974) and may also play a role in the intrinsic pathway of blood coagulation (Weupper, 1973; Saito et al., 1974; Weiss et al., 1974) and fibrinolysis (Laake & Vennerod, 1974; Ogston et al., 1969; Mandle & Kaplan, 1977; Ichinose et al., 1986). Plasma prekallikrein ($M_r$ 85 000) is converted to plasma kallikrein by factor XIIa by the cleavage of a single internal Arg-Ile peptide bond (Griffin & Cochrane, 1979; Heimark et al., 1980). Plasma kallikrein is a serine protease that is composed of a heavy chain (371 amino acids) and a light chain (248 amino acids) held together by a disulfide bond(s) (Chung et al., 1986).

The amino acid sequence of plasma prekallikrein is 58% identical with factor XI, a plasma protein that is converted to factor XIa in the early phase of the blood coagulation cascade. The amino-terminal regions of both plasma prekallikrein and factor XI also contain 4 tandem repeats of 90 (or 91) residues that show a high degree of amino acid sequence identity (Chung et al., 1986; Fujikawa et al., 1986). Furthermore, each of the four tandem repeats contains six half-cystine (Cys)[1] residues that are highly conserved, being present in the same position in each of the four tandem repeats in both proteins. Also, the positions of the six Cys residues in the two proteins do not align with the Cys residues of any known protein domain containing disulfide bonds. These data suggest that plasma prekallikrein as well as factor XI may have three unique disulfide bonds in each of their four tandem repeats. Plasma prekallikrein contains two additional Cys residues in the fourth repeat, while each of the first and fourth

† This work was supported in part by Grant HL 16919 from the National Institutes of Health. A preliminary report was presented at the Fourth Annual Symposium of the Protein Society held in San Diego, CA, Aug 11-15, 1990.

[1] Abbreviations: Cys, half-cystine; PTH, phenylthiohydantoin.

0006-2960/91/0430-2050$02.50/0 