Site-Specific Modification of Pre-mRNA: The 2′-Hydroxyl Groups at the Splice Sites

Melissa J. Moore and Phillip A. Sharp

A simple and efficient method for synthesizing long, site-specifically modified RNA molecules was developed whereby segments of RNA were joined with the use of bacteriophage T4 DNA ligase. A single hydrogen or *O*-methyl group was substituted for the 2′-hydroxyl group at either splice site of a nuclear pre-messenger RNA substrate. Splicing of the modified pre-messenger RNA's in vitro revealed that, although a 2′-hydroxyl is not absolutely required at either splice site, the 2′-hydroxyl at the 3′ splice site is important for the second step of splicing. These results are compared to previous studies of analogous 2′-hydroxyl groups in the self-splicing *Tetrahymena* group I intron.

The precise removal of intervening sequences, or introns, by the process called splicing is a fundamental step in the maturation of most eukaryotic pre-mRNA's. Excision of these sequences takes place within the spliceosome, a 50S to 60S complex composed of the pre-mRNA, four small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U4/6, and U5) and an as yet undetermined number of associated protein factors (*1*). Within the spliceosome, splicing proceeds by way of a two-step mechanism involving sequential transesterification reactions. The first step (Fig. 1, left) entails cleavage at the 5′ splice site with concurrent formation of a 2′,5′-phosphodiester bond between the first nucleotide of the intron and an A residue (the branch site) located 18 to 100 nucleotides (nt) upstream of the 3′ splice site. The second step (Fig. 1, right) takes place at the 3′ splice site and results in ligation of the two exons and release of the intron as a lariat. Because the self-catalyzed excision of group II introns proceeds through the same two-step pathway (*2*), the two intron types are thought to be evolutionarily related (*3*). Group I self-splicing is similar except that the 3′-hydroxyl (-OH) of a free guanyl nucleotide is the nucleophile for the first step (resulting in a 3′,5′-phosphodiester bond) (*2*). Of the three intron types, group I introns are the best characterized chemically, serving as the paradigm for catalytic RNA's (*4*). Because of the complexity of the spliceosome and the size of the pre-mRNA substrates, however, nuclear pre-mRNA splicing has so far proven recalcitrant to more detailed biochemical analysis. Therefore, little is known about either the exact chemical mechanisms of catalysis in this system or the nature of the participating spliceosomal groups.

The authors are in the Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

We now describe a method for studying cellular processes, including nuclear pre-mRNA splicing, that involve long RNA molecules. In this method for synthesizing long RNA's containing internal site-specific modifications, the RNA is made in segments, one of which contains a desired modification. The segments are then joined by ligation with bacteriophage T4 DNA ligase in conjunction with a bridging oligodeoxynucleotide template. This method was used to study the roles in splicing of the 2′-OH groups (Fig. 1) immediately adjacent to the phosphates at the 5′ and 3′ splice sites of a nuclear pre-mRNA substrate by specific chemical replacement with both 2′-deoxy (-H) and 2′-O-methyl (-OCH_3) groups.

Specific ligation of RNA molecules. Because of problems inherent in selectively protecting the 2′-OH moiety while activating the 3′-OH group for phosphoramidite coupling, routine chemical synthesis of RNA has been limited to 10- to 20-nt oligomers. Longer RNA's containing internal, site-specific chemical modifications have been prepared with the RNA ligase of T4 bacteriophage (*5*), but this enzyme's utility is limited by its high Michaelis constant (K_m) for polynucleotides (*6*) and its discrimination against certain acceptor and donor sequences (*7*). In addition, T4 RNA ligase can produce significant amounts of undesirable side products, including circular and oligomeric RNA molecules, unless the 3′-OH terminus of the phosphate donor RNA is protected or destroyed (*6*). A chemical ligation of RNA with carbodiimide or cyanogen bromide activated phosphodiesters (*8*) overcomes many of these problems, but the method has not become general, partly because of its chemical incompatibility with certain nucleotide modifications.

The DNA ligase of T4 bacteriophage has an RNA ligase activity. In the early 1970's, it was demonstrated that the enzyme (then known as T4 polynucleotide ligase) could efficiently ligate oligoribonucleotide homopolymers in the presence of complementary deoxyhomopolymer templates (*9*). Yet, to our knowledge, this activity has not been widely used to join RNA's of defined sequence. This is surpris-

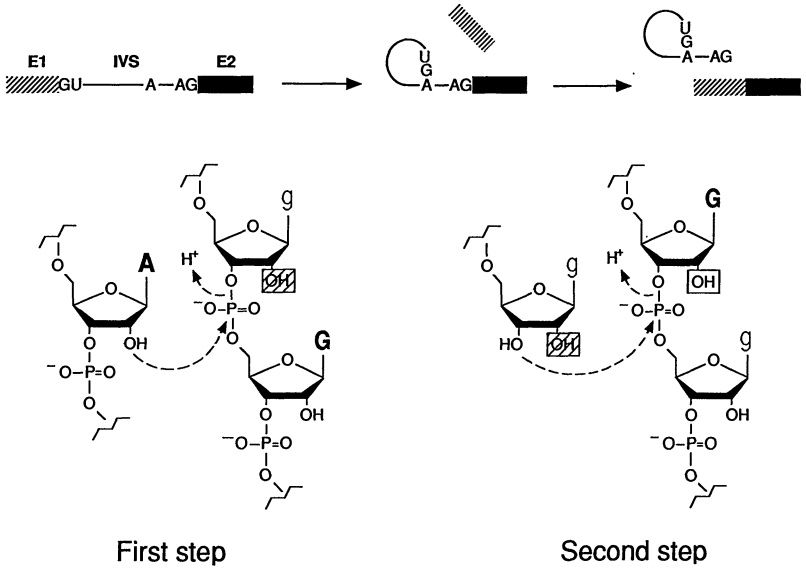

Fig. 1. The two steps of nuclear pre-mRNA splicing. (**Upper**) Schematic representation of the two steps showing the substrate (E1-IVS-E2), intermediates (E1 and IVS-E2), products (IVS and E1-E2), and conserved nucleotides. E1, 5′ exon; IVS, intervening sequence; E2, 3′ exon. (**Lower**) The chemistry of the two steps showing the 2′-OH groups adjacent to the phosphates at the 5′ (striped box) and 3′ (open box) splice sites. Conserved nucleotides are shown as upper case. Dashed arrows indicate only the overall chemical outcome; whether the chemistry occurs through a direct or indirect mechanism has not been determined.

ing since we have observed that, in the presence of a complementary deoxyoligonucleotide (cDNA) bridge, T4 DNA ligase can efficiently join long RNA molecules in a highly selective manner.

Representative ligation reactions of the RNA fragments used in this study and their cDNA bridges (Fig. 2A) illustrate several key features of RNA-RNA ligations catalyzed by T4 DNA ligase (*10*). The low K_m of T4 DNA ligase for double-stranded polynucleotides (from 10^{-8} to 10^{-7} M) (*11*) permits efficient ligations to be performed at submicromolar to micromolar concentrations of RNA. This concentration range is readily attainable and is ideal for the synthesis of high specific activity ^{32}P-labeled RNA's, where only femtomoles to picomoles of product are required. For example (Fig. 2B, lane 4), T4 DNA ligase (4 units) converted 45 percent of RNA(55–236) (1 μM) into the desired product, RNA(1–236), in the presence of RNA(1–54) (3 μM) and cDNA(68–44) (2 μM). In contrast, T4 RNA ligase prefers single-stranded polynucleotides at concentrations in the millimolar range (*12*). Under conditions similar to those above, T4 RNA ligase (1.6 units) yielded no specific ligation product in a reaction containing RNA(1–54) (3 μM) and RNA(55–236) (1 μM) (Fig. 2B, lane 7).

Since T4 DNA ligase will only ligate junctions in double-stranded regions (Fig. 2B, lanes 4 and 6), there is little tendency to circularize or oligomerize the phosphate donor RNA. In addition, the cDNA template confers significant sequence specificity on the reaction. This is important when RNA's synthesized with either T7, T3, or SP6 RNA polymerase are used, because these RNA's often contain a high percentage of products ending with extra nontemplated nucleotides at the 3′ terminus [N+1 products (*13*)]. Since the T4 RNA ligase reaction is not template directed, it cannot distinguish N and N+1 acceptor sequences. Therefore, a high percentage of such ligations incorporate these extra nonencoded nucleotides (Fig. 2C, upper). In contrast, because T4 DNA ligase prefers precise base-pairing at the ligation junction (*14*), it is highly selective against extra bases and only ligates acceptor molecules having the correct 3′-terminal nucleotide (Fig. 2C, lower). Thus, after correction for an approximately 50 percent content of N+1 runoff transcription products in RNA (1–54) (*15*), the true ligation efficiency for the accurately bridged RNA's in Fig. 2B approaches 90 percent with T4 DNA ligase.

Synthesis and splicing of pre-mRNA's with site-specific modifications. The T4 DNA ligase was used to incorporate nucleotides with either 2′-H or 2′-OCH_3 modifications immediately adjacent to the phosphates at the 5′ and 3′ splice sites of a pre-mRNA substrate, RNA(1–236) (*16*). RNA(1–236) (E1-IVS-E2) comprises a 55-nt 5′ exon (E1) and a 53-nt 3′ exon (E2) interrupted by a 128-nt intron (IVS). The IVS contains exact consensus sequences at the branch site and both splice sites, as well as an uninterrupted 20-nt polypyrimidine tract. This substrate is spliced efficiently in vitro. Modified nucleotides were incorporated by first chemically synthesizing GpG dinucleotides in which the ribose of the 5′ guanosine was either 2′-H or 2′-OCH_3, while the 3′ guanosine was unmodified (resulting in dG_prG or mG_prG) (*17*). These dinucleotides were used to prime transcription by T7 RNA polymerase to generate an RNA molecule [RNA(55–236) or RNA(183–236)] (*16*), which was then 5′-phosphorylated and ligated to the 3′ end of a second transcript [RNA(1–54) or RNA(1–182), respectively] (*16*) such that the original GpG dinucleotide spanned the desired splice site (Fig. 2A). Unmodified (all-ribo) control RNA's were generated in parallel with the use of rG_prG dinucleotide. The two modified dinucleotides, dG_prG and mG_prG, were incorporated with efficiencies similar to rG_prG, as indicated by digestion of the ligation products with T1 ribonuclease (*15*). T1 ribonuclease cleaves only after guanosine residues, but it will not cleave after either 2′-deoxy- or 2′-*O*-methylguanosine (*18*). The E1-IVS-E2 RNA's were either uniformly labeled by inclusion of [α-^{32}P]UTP in the transcription reactions or by phosphorylation of the 3′ ligation substrate with polynu-

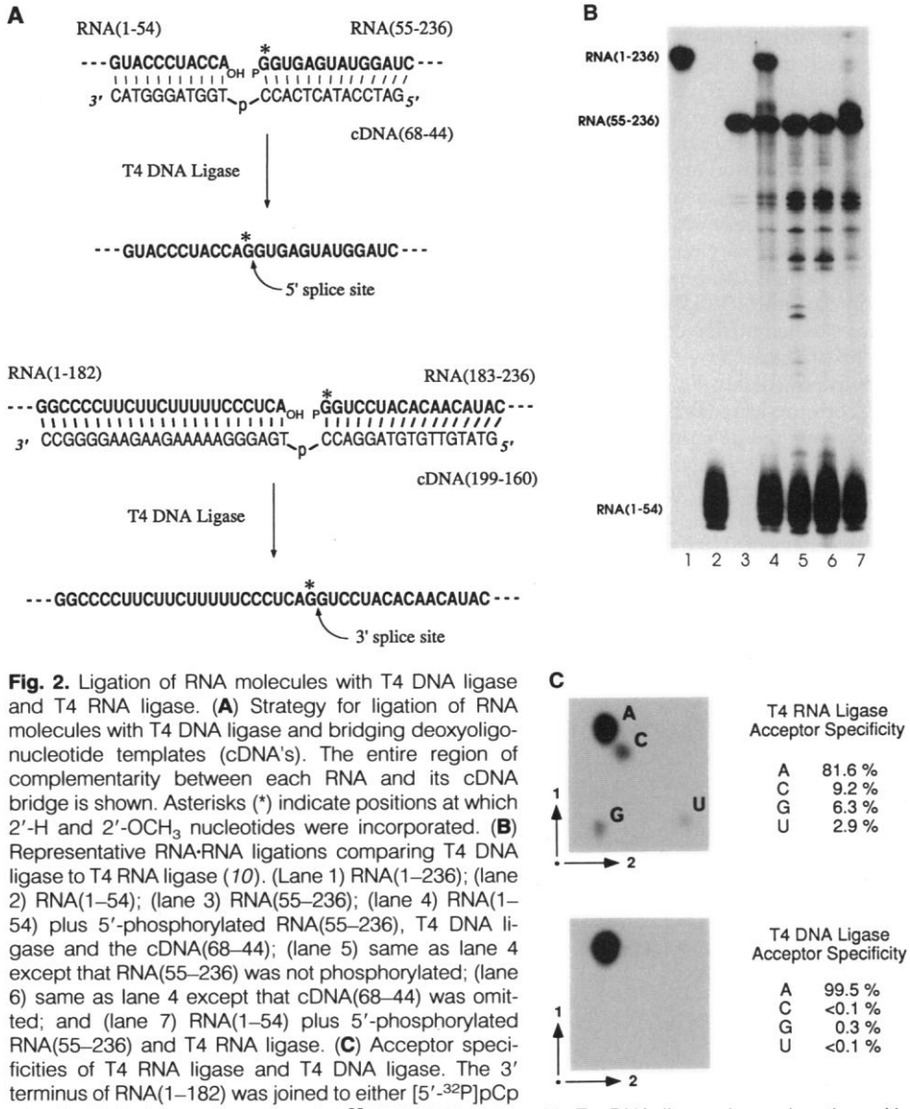

Fig. 2. Ligation of RNA molecules with T4 DNA ligase and T4 RNA ligase. (**A**) Strategy for ligation of RNA molecules with T4 DNA ligase and bridging deoxyoligonucleotide templates (cDNA's). The entire region of complementarity between each RNA and its cDNA bridge is shown. Asterisks (*) indicate positions at which 2′-H and 2′-OCH_3 nucleotides were incorporated. (**B**) Representative RNA·RNA ligations comparing T4 DNA ligase to T4 RNA ligase (*10*). (Lane 1) RNA(1–236); (lane 2) RNA(1–54); (lane 3) RNA(55–236); (lane 4) RNA(1–54) plus 5′-phosphorylated RNA(55–236), T4 DNA ligase and the cDNA(68–44); (lane 5) same as lane 4 except that RNA(55–236) was not phosphorylated; (lane 6) same as lane 4 except that cDNA(68–44) was omitted; and (lane 7) RNA(1–54) plus 5′-phosphorylated RNA(55–236) and T4 RNA ligase. (**C**) Acceptor specificities of T4 RNA ligase and T4 DNA ligase. The 3′ terminus of RNA(1–182) was joined to either [5′-^{32}P]pCp with T4 DNA ligase (upper) or [5′-^{32}P]RNA(183–236) with T4 DNA ligase in conjunction with cDNA(199–160) (lower). Ligation products were purified by gel electrophoresis and then digested with T2 ribonuclease; the resultant nucleoside 3′-monophosphates were separated by two-dimensional thin-layer chromatography (*37*). Percentages given are for the relative amount of label found at each nucleotide migration position as indicated.

cleotide kinase and [γ-^{32}P]ATP (adenosine triphosphate). The latter method yielded E1-IVS-E2 molecules containing a single ^{32}P-labeled phosphate immediately 5′ to the modified nucleotide (as in Fig. 5).

The effects of the ribose modifications on E1-IVS-E2 splicing were determined by following the generation of intermediate and product RNA's during in vitro splicing reactions (Fig. 3) (*19*). The all-ribo E1-IVS-E2 RNA's produced by joining two RNA molecules at either the 5′ or 3′ splice site were as active for splicing as all-ribo E1-IVS-E2 RNA synthesized as a single runoff transcript (*15*). This further confirmed the specificity of the T4 DNA ligase joining reaction.

Modifications at the 5′ splice site. The 2′-OH group at the 5′ splice site could potentially influence either or both steps of splicing. In the first step, this 2′-OH is adjacent to the 3′-OH that is ultimately displaced from the phosphate at the 5′ splice site by the 2′-OH of the branch site adenosine (Fig. 1); in the second step, the same 3′-OH must act as a nucleophile to effect exon ligation at the 3′ splice site. In either step, the adjacent 2′-OH might function in a number of ways (see below) and its function could vary between the two steps.

Surprisingly, substitution of a 2′-H group for the 2′-OH adjacent to the 5′ splice site had no detectable effect on either the first or second step of splicing (Fig. 3); both intermediates (E1 and IVS-E2 RNA's) and both products (IVS and E1-E2 RNA's) were generated at rates similar to those of the all-ribo intermediates and products (Table 1). In contrast, substitution of a 2′-OCH_3 group for the same 2′-OH affected the two steps of splicing differently (Fig. 3, lanes 19 to 27). The rate of the first step, as measured by production of E1 and IVS-E2 RNA's, was similar to that of all-ribo RNA (Table 1). An effect, however, was observed for the second step of splicing, as evidenced by the accumulation of the E1 and IVS-E2 RNA intermediates and the dramatically decreased rate of appearance of the E1-E2 RNA product (15 times lower).

Both pre-mRNA substrates containing modifications at the 5′ splice site yielded a cleaved E1 intermediate RNA with an increased electrophoretic mobility relative to all-ribo E1 RNA (Fig. 4A, lanes 1 to 3). The degree of increase was equivalent to that expected for an RNA that was shorter by one nucleotide. To examine whether the site of cleavage was correct, we synthesized E1-IVS-E2 RNA's that contained a single ^{32}P-labeled phosphate immediately 5′ to the modified nucleotide at the 5′ splice site (Fig. 5, lower left). If the site of cleavage of the modified E1-IVS-E2 RNA's was shifted one nucleotide in the 5′ direction relative to that of all-ribo E1-IVS-E2 RNA, then the labeled phosphate between the A and G residues would reside in the IVS-E2 and IVS RNA's, and not in the E1 RNA. However, all of the label was found in the cleaved E1 (Fig. 4A, lanes 4 to 6) and ligated E1-E2 RNA's (*15*). In addition, each singly labeled E1 RNA comigrated with its uniformly labeled counterpart (compare lanes 4 to 6 to lanes 1 to 3 in Fig. 4A). Together, this suggested that cleavage had occurred at the normal 5′ splice site. Absolute confirmation came when the purified, singly-labeled all-ribo, 2′-H, and 2′-OCH_3 E1 RNA's were digested by RNase A (Fig. 5, lanes 15 to 17). All of the singly-labeled E1 RNA's had the expected -ApG_{OH} 3′-terminus. Furthermore, digestion of the all-ribo and 2′-H E1-E2 product RNA's with RNase A and T1 ribonuclease (Fig. 5, lanes 18 and 19) confirmed that the E1 RNA ending in 2′-H guanosine had been precisely used for the second step of splicing. Therefore, the observed increases in electrophoretic mobility of the 2′-H and 2′-OCH_3 E1 RNA's must have resulted from the chemical modifications present at their 3′-termini and not from a difference in overall length of the RNA's (*20*).

Modifications at the 3′ splice site. Neither modification introduced at the 3′ splice site affected the first step of splicing (Fig. 3, lanes 28 to 45); for both the 2′-H and 2′-OCH_3 modified substrates, the E1 and IVS-E2 RNA intermediates were generated with similar initial kinetics to those of the all-ribo substrate (Table 1). The presence of the modifications in the lariat IVS-E2 RNA's was confirmed by RNase A and T1 ribonuclease digestion of IVS-E2 RNA's generated from E1-IVS-E2 RNA substrates that were singly labeled immediately 5′ to the 3′ splice site (Fig. 5, lower right and lanes 20 to 22). Introduction of a 2′-H group at the 3′ splice site, however, did dramatically alter the rate of the second step of splicing; the rate of appearance of the E1-E2 RNA was 19 times lower (Fig. 3, lanes 28 to 36, and Table 1). Substitution of a 2′-OCH_3 group at the 3′ splice site was also detrimental, but less so, with the production of E1-E2 RNA reduced to one-seventh the original rate (Fig. 3, lanes 37 to 45, and Table 1).

In all three cases where modification of the substrate RNA reduced the rate of the second step of splicing (2′-OCH_3 at the 5′ splice site and 2′-H or 2′-OCH_3 at the 3′ splice site), RNA's migrating as excised lariat IVS RNA accumulated at a much higher rate than E1-E2 RNA (Fig. 3, lanes 19 to 45) (*21*). Yet, since both IVS and E1-E2 RNA's are products of the second step, they should have been generated in equimolar amounts. One possible explanation is that the modifications effected an increase in nucleolytic hydrolysis at the 3′ splice site, thereby partially uncoupling intron excision from exon ligation (*22*). This would be expected to produce lariat IVS RNA and free E2 RNA. The latter would migrate in

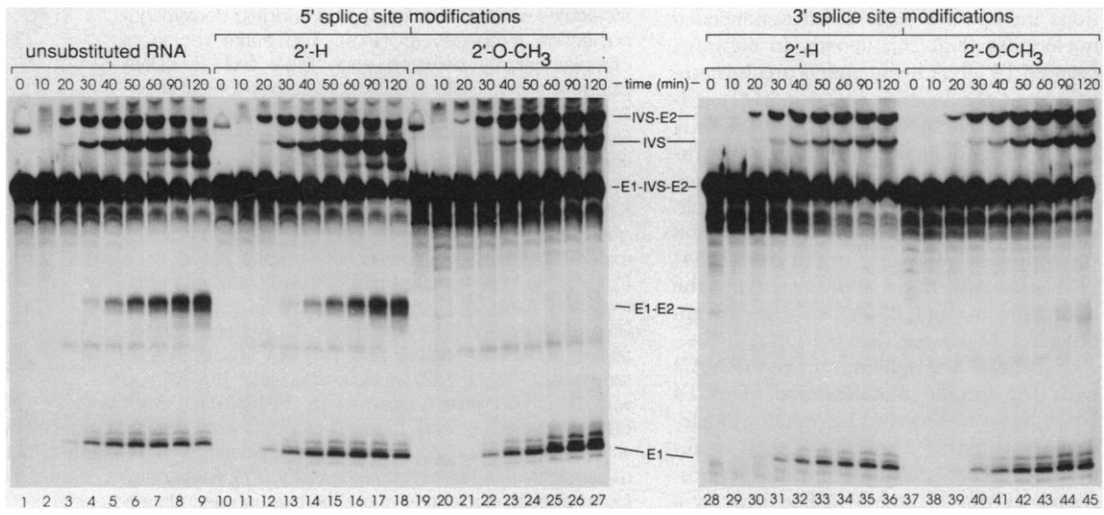

Fig. 3. Splicing time courses for all-ribo and site-specifically modified E1-IVS-E2 RNA's. Uniformly labeled E1-IVS-E2 RNA's were incubated under splicing conditions (*19*) for the times indicated, and then separated on a denaturing 15 percent polyacrylamide gel. The electrophoresis time for the right panel was somewhat longer than that for the left panel. E1-E2 product RNA's were visible in the 5′ splice site-(2′-OCH_3) and 3′ splice site-(2′-H) time courses upon longer exposure.

close proximity to the excised E1 RNA intermediate, making its detection difficult in reactions such as those shown in Fig. 3. This possibility was ruled out, however, when we examined the splicing of an E1-IVS-E2 substrate RNA that contained the 2′-H modification at the 3′ splice site and the ^{32}P label only in the E2 exon. The absence of detectable free E2 RNA in these reactions (*15*) demonstrated that increased hydrolysis could not account for the differential rates of appearance of IVS and E1-E2 RNA's.

A more probable explanation for the above paradox is that some of the RNA molecules migrating at the position of IVS RNA in the 15 percent polyacrylamide gels (Fig. 3) were generated by nucleolytic degradation of the accumulated IVS-E2 RNA. This became evident when the splicing products were separated on a 6 percent polyacrylamide gel, which permitted resolution of the lariat RNA's according to their length (Fig. 4B). Singly labeled all-ribo and 3′ splice site modified IVS RNA's resolved into a family of closely migrating species. The fastest migrating member of this family (α in Fig. 4B) consisted of true IVS RNA having a 3′ terminus of . . .ApG_{OH}, as shown by ribonuclease (RNase) A digestion (Fig. 5, lanes 23 to 26 and 31 to 32). The slower migrating species (β, γ) are lariat RNA's whose 3′-termini extend one or more nucleotides into the sequence of the E2 exon (Fig. 5, lanes 27 to 30). These slower migrating species were present in both the control all-ribo sample (Fig. 4B, lane 1; Fig. 5, lanes 23 and 24) and the sample RNA's with modifications at the 3′ splice site (Fig. 4B, lanes 2 and 3; Fig. 5, lanes 27 to 32). Therefore, their presence was not a function of chemical modification of the splice sites. The longer IVS-type RNA's were probably generated by nucleolytic digestion of the IVS-E2 RNA's accumulated in spliceosomes, producing the equivalent of a "footprint" on the IVS-E2 lariat. Similar footprinting has been described (*23*). This is consistent with the observation that the ratio of the longer IVS-type molecules (β, γ) to true IVS molecules (α) could be reduced by the addition of tRNA as a competitor for the nucleases in the splicing reactions (*15*). In addition, the 2′-H and 2′-OCH_3 substitutions at the 3′ splice site reduced the rate of accumulation of true IVS molecules (as analyzed on low percentage gels) (*15*) to the

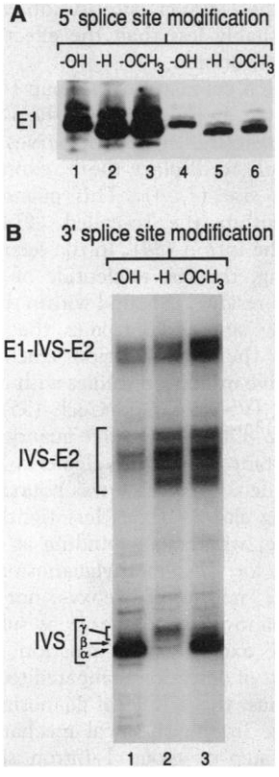

Fig. 4. Electrophoretic analysis of splicing intermediates and products. (**A**) Side-by-side mobility comparison of all-ribo and 5′ splice site modified E1 RNA's in a denaturing 15 percent polyacrylamide gel. (Lanes 1 to 3) Uniformly labeled RNA's; (lanes 4 to 6) RNA's singly labeled adjacent to the 5′ splice site. (**B**) Separation of lariat IVS RNA's by electrophoresis in a denaturing 6 percent polyacrylamide gel. All lanes contain RNA's that were singly labeled adjacent to the 3′ splice site. Arrows (α, β, γ) indicate the particular IVS species that were purified and separately analyzed by ribonuclease digestion in Fig. 5. All samples were incubated under splicing conditions (*19*) for 120 minutes prior to electrophoresis.

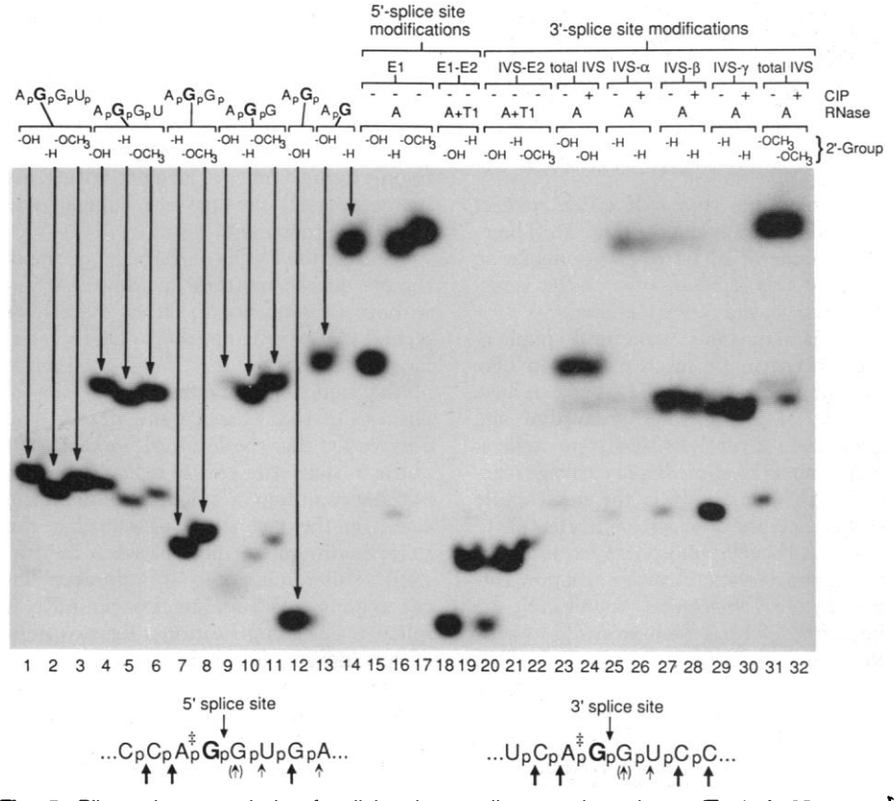

Fig. 5. Ribonuclease analysis of splicing intermediates and products. (**Top**) A 25 percent polyacrylamide gel of singly labeled oligonucleotides generated from all-ribo and splice site modified RNA's. (Lanes 1 to 14) Control RNA's (arrows) having the nucleotide and phosphate composition indicated (*39*). The group (-OH, -H, or -OCH_3) at the 2′ position of the boldface **G** residue is indicated immediately above each lane. (Lanes 15 to 32) Ribonuclease and phosphatase analysis of singly labeled splicing intermediates and products. E1 (lanes 15 to 17) and E1-E2 (lanes 18 to 19) RNA's were generated by incubation of the appropriately modified (2′-OH, 2′-H, or 2′-OCH_3) E1-IVS-E2 RNA labeled at the 5′ splice site under splicing conditions (*19*) for 120 minutes, and then purified on a 15 percent polyacrylamide gel. The IVS-E2 (lanes 20 to 22) and IVS (lanes 23 to 32) RNA's labeled at the 3′ splice site were prepared similarly (15 percent gel) except that the splicing reactions were for 170 minutes. The total 2′-H IVS RNA was further purified by recovery from a 6 percent polyacrylamide gel (Fig. 4B) to yield IVS-α (lanes 25 to 26), IVS-β (lanes 27 to 28), and IVS-γ (lanes 29 to 30). Samples that were treated with calf intestinal phosphatase (CIP) to remove 3′-phosphates are indicated (+). (**Bottom**) Sequences around the 5′ and 3′ splice sites showing RNase A (bold arrows) and T1 ribonuclease (small arrows) cleavage sites. The small arrow in parentheses denotes the cleavage site that was not cut by T1 when either modification (2′-H or 2′-OCH_3) was present on the boldface **G** residue. All RNA's contained a single labeled phosphate (‡) adjacent to either the 5′ or 3′ splice site.

Table 1. Relative rates of the first and second steps of splicing using site specifically modified pre-mRNA substrates. Polyacrylamide gels were quantified (Molecular Dynamics PhosphorImager) and the relative amount of RNA in each band was expressed as a percentage of the total obtained by summing the values for E1-IVS-E2, E1, IVS-E2, IVS, and E1-E2 at that time point (*36*). Initial rates were chosen from the linear portion of each curve (at least three time points) and normalized to the respective all-ribo rate. Though the rates given are for the time-courses shown in Fig. 3, the results were reproducible in other experiments.

	First step		Second step	
	E1	IVS-E2	IVS	E1-E2
All-ribo	1.0	1.0	1.0	1.0
5′ splice site				
2′-H	1.2	1.0	0.75	0.88
2′-OCH_3	1.1	1.1	*	0.07
3′ splice site				
2′-H	1.3	1.3	*	0.05
2′-OCH_3	1.1	1.2	*	0.14

*For these modified RNA's, the rate of accumulation of IVS-like species was not representative of the rate of the second step of splicing.

same extent that they reduced the rate of E1-E2 RNA accumulation (Fig. 3). Therefore, the rate of E1-E2 RNA accumulation (Table 1) was representative of the actual rate of the second step of splicing.

Mechanistic and structural implications. In enzymatic reactions that involve RNA substrates, the 2′-OH group can have a number of distinct roles. In alkaline and ribonuclease hydrolysis (*24*), as well as RNA hammerhead-mediated cleavage reactions (*25*), the 2′-OH is the nucleophile that displaces the downstream nucleotide to form a 2′,3′-cyclic phosphate terminus. In the endonucleolytic cleavage reaction catalyzed by the *Tetrahymena* group I IVS, the adjacent 2′-OH has been proposed to activate the neighboring 3′-oxygen leaving group (*26*). The 2′-OH group can also serve as a recognition element by either accepting or donating hydrogen bonds (*27*) or simply ensuring a particular conformation of the ribose ring (*27, 28*). In addition, 2′-OH's might function as ligands for the coordination of metal cations (*29*).

Substitution of the 2′-OH at the 5′ splice site with a 2′-H had no effect on the overall rate of 5′ splice site cleavage as compared to the control RNA. Thus, neither formation of the spliceosome nor execution of the first chemical step, to form the 2′,5′ branch, involves recognition or chemistry that absolutely requires the 2′-OH on the adjacent ribose ring. This same 2′-OH is also not required for the second step of splicing, as the 2′-H modification again produced no decrease in the rate of that step when compared to the control RNA. These results must be interpreted with caution, however, as the nature of the rate limiting step for pre-mRNA splicing is not known. Therefore, it is possible that the 2′-H modification could be having effects on the actual chemical reactions without those reactions becoming rate limiting on our assay.

The above results can be compared to those obtained for the endonucleolytic cleavage reaction catalyzed by the *Tetrahymena* group I IVS. Using an all-deoxy substrate, Herschlag and Cech (*26*) showed that removal of the 2′-OH adjacent to the site of cleavage (analogous to the 5′ splice site) resulted in a large (more than four orders of magnitude) decrease in the rate of the chemical reaction at that site. (This suggested that in group I introns, the 2′-OH at the 5′ splice site is directly involved in accelerating the initial cleavage reaction, possibly by facilitating protonation of the adjacent 3′ oxygen atom to make it a better leaving group.) Yet, because product release is normally rate limiting for catalytic turnover in the *Tetrahymena* system, the large decrease in the rate of the chemical reaction with the deoxy substrate gave rise to only a single order of magnitude decrease in the coverall cleavage rate compared to an all-ribo substrate.

Because the 2′-OH at the 5′ splice site of the pre-mRNA substrate could be removed without consequence to either the first or second step of splicing, the 2′-OCH_3 modification at that site served as a good probe of the steric environment around the 2′ position of the 5′ splice site ribose. The observation that the 2′-OCH_3 modification at the 5′ splice site greatly reduced the rate of the second step of splicing, but had little effect on the first step, suggests that the steric environments differ between the two steps. This result points to a contrast between group I IVS and nuclear pre-mRNA splicing. For group I introns, the two steps of splicing can be explained as identical forward and reverse reactions that are catalyzed by a single reactive center (*4, 30*). The two steps of nuclear pre-mRNA splicing, however, must be distinct as the substrates for the two reactions are markedly different. The chemistry at the 5′ splice site involves an incoming 2′ oxygen atom, which forms a 2′,5′-phosphodiester bond, whereas that at the 3′ splice site utilizes a 3′ oxygen to form a 3′,5′-phosphodiester bond. It is possible that the two reactions could share most constituents of a single catalytic center, or, alternatively, there could be two distinct active sites. The fact that the 2′-OCH_3 modification at the 5′ splice site differentially affected the rates of the two steps of splicing is consistent with the two processes differing at least in steric detail.

Introduction of either 2′-H or 2′-OCH_3 at the 3′ splice site had no effect on the rate of the first step of splicing. This suggests that the chemical structure of the ribose adjacent to the phosphate at the 3′ splice site is not recognized in formation of the spliceosome and cleavage at the 5′ splice site. This is not surprising, since the first step of splicing can occur in the total absence of a 3′ splice site in both yeast (*31*) and mammals (*32, 33*).

Although they had no effect on the first step, both modifications at the 3′ splice site did reduce the rate of the second step of splicing. When the 2′-OH adjacent to the 3′ splice site was replaced by 2′-H, the rate of the second step was 19 times lower, while the 2′-OCH_3 modification produced a rate that was only 7 times lower. It is difficult to surmise, however, whether these reductions in rate indicate a direct role for this 2′-OH group in the chemical mechanism of exon ligation rather than an indirect role as a recognition element. Perhaps a specific role for the oxygen atom is suggested by the effect of the 2′-OCH_3 substitution, which is distinguishably less than the effect of the 2′-H substitution.

Again, a comparison to group I introns is warranted. In those introns, the 3′-OH of a free guanosine nucleotide serves as the nucleophile to displace the 5′ exon at the 5′ splice site (*2, 4*). This guanosine is bound within the so-called "G-binding site" of the intron (*34*). In the second step of splicing, the last nucleotide of the intron, a G residue, is bound within the same G-binding site and becomes the leaving group in the exon ligation reaction. In competitive inhibition studies with the *Tetrahymena* IVS, Bass and Cech (*35*) found that the 2′-OH group of free guanosine was an important recognition element. Specifically, 2′-deoxyguanosine was bound in the G-binding site 34 times less tightly than guanosine, whereas no binding at all was observed for 2′-O-methylguanosine. Furthermore, neither 2′-deoxy- nor 2′-O-methylguanosine could serve as substrates for the 5′ exon cleavage reaction (with a 10^{-5} limit of detection compared to guanosine). Thus, the 2′-OH of guanosine must participate in the chemical mechanism of the first step of group I intron splicing. Hence, the 2′-OH adjacent to the 3′ splice site of group I introns may also prove to be essential because it is recognized by the same reactive center. Proof of these propositions, along with our data, would constitute strong evidence that there are fundamental chemical differences between the splicing of group I and pre-mRNA introns. We have clearly shown that neither 2′-OH group adjacent to a splice site in pre-mRNA's is absolutely required for splicing.

The simple technique described for introducing chemical modifications or a radioactive label (or both) at specific positions of polyribonucleotides should prove

useful in the study of many processes that involve long RNA molecules, including ribosomal translation, RNA transport and localization, and RNA catalysis. Our initial studies have already provided key insights into the reactions that are catalyzed by the spliceosome.

REFERENCES AND NOTES

1. P. A. Sharp, *Science* **235**, 766 (1987); T. Maniatis and R. Reed, *Nature* **325**, 673 (1987); C. Guthrie, *Science* **253**, 157 (1991); S. W. Ruby and J. Abelson, *Trends Genet.* **7**, 79 (1991).
2. P. A. Sharp, *Cell* **42**, 397 (1985); T. R. Cech, *ibid.* **43**, 713 (1985).
3. T. R. Cech, *ibid.* **44**, 207 (1986); A. Jacquier, *Trends Biochem. Sci.* **15**, 351 (1990); P. A. Sharp, *Science* **254**, 663 (1991).
4. T. R. Cech, *Science* **236**, 1532 (1987).
5. W. L. Wittenberg and O. C. Uhlenbeck, *Biochemistry* **24**, 2702 (1985).
6. O. C. Uhlenbeck and R. I. Gumport, *The Enzymes* **15**, 31 (1982); P. J. Romaniuk and O. C. Uhlenbeck, *Methods Enzymol.* **100**, 52 (1983).
7. T. E. England and O. C. Uhlenbeck, *Biochemistry* **17**, 2069 (1978); E. Romaniuk, L. W. McLaughlin, T. Neilson, P. J. Romaniuk, *Eur. J. Biochem.* **125**, 639 (1982).
8. Z. Shabarova, *Biochimie* **70**, 1323 (1988).
9. K. Kleppe, J. H. van de Sande, H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 68 (1970); G. C. Fareed, E. M. Wilt, C. C. Richardson, *J. Biol. Chem.* **246**, 925 (1971); H. Sano and G. Feix, *Biochemistry* **13**, 5110 (1974).
10. Except as noted in the legend of Fig. 2B, ligation conditions were: 50 mM tris-HCl (pH 7.5), 10 mM $MgCl_2$, 20 mM DTT (dithiothreitol), 1 mM ATP, BSA (bovine serum albumin) at 50 μg/ml, 3 μM RNA(1–54), 2 μM cDNA(68–44), 1 μM RNA(55–236), T4 DNA ligase [United States Biochemical (USB) or New England Biolabs (NEB)] at 1 U/μl for 4 hours at 25°C. The T4 RNA ligase reaction was the same, except that the cDNA(68–44) was omitted and T4 RNA ligase (NEB) at 0.4 U/μl was substituted for T4 DNA ligase. The T4 DNA ligase enzyme units were as defined by USB and by B. Weiss *et al.* [*J. Biol. Chem.* **243**, 4543 (1968)]; T4 RNA ligase enzyme units were defined by NEB and by T. England, R. Gumport, and O. Uhlenbeck [*Proc. Natl. Acad. Sci. U.S.A.* **74**, 4839 (1977)] and by A. Sugino *et al.* [*J. Biol. Chem.* **252**, 3987 (1977)]. Prior to the ligation, G_pG-primed RNA's (*16*) were 5′-phosphorylated with ATP and T4 polynucleotide kinase (NEB). In DNA ligase reactions, the two RNA's to be joined were hybridized to the appropriate cDNA template by heating a mixture of the three to 92°C for 1 minute and then holding for 5 minutes at ambient temperature. Preparative scale ligation reactions were done similarly, with RNA substrate concentrations ranging from 0.5 to 3 μM and T4 DNA ligase concentrations up to 2 U/μl.
11. M. J. Engler and C. C. Richardson, *Enzymes* **15**, 3 (1982).
12. The K_m of T4 RNA ligase for polynucleotides has not been determined, but it may be in excess of 1 mM [O. C. Uhlenbeck and V. Cameron, *Nucleic Acids Res.* **4**, 85 (1977)].
13. For references see J. K. Yisraeli and D. A. Melton, *Methods Enzymol.* **180**, 42 (1989); J. F. Milligan and O. C. Uhlenbeck, *ibid.*, p. 51.
14. While precise base pairing at the ligation junction is optimal for DNA-DNA ligations by T4 DNA ligase, ligation of sequences containing nucleotide mismatches has been observed [C. M. Tsiapalis and S. A. Narang, *Biochem. Biophys. Res. Commun.* **39**, 631 (1970); V. Sgaramella and H. G. Khorana, *J. Mol. Biol.* **72**, 427 (1972); B. Weiss, *ibid.* **103**, 669 (1976)].
15. M. J. Moore and P. A. Sharp, data not shown.
16. pPIP85.A is an optimized splicing construct derived from pPIP2 [M. G. Garcia-Blanco, S. F. Jamison, P. A. Sharp, *Genes Dev.* **3**, 1874 (1989)] that contains the following 237-nt sequence between the T7 promoter and Hind III site of pBS⁻ (Stratagene):

5′-GGGCGAATTCGAGCTCACTCTCTTCCGCATCGCTGTCTGCGAGG
TACCCTACCAG↓GTGAGTATGGATCCCTCTAAAAGCGGGCATGACT
TCTAGAGTAGTCCAGGGTTTCCGAGGGTTTCCGTCGACGATGTCAG
CTCGTCTCGAGGGCGTACTAACTGGGCCCCTTCTTCTTTTTCCCTCA
G↓GTCCTACACAACATACTGCAGGACAAACTCTTCGCGGTCTCTGC
ATGCAAGCTT-3′

Arrows (↓) indicate the 5′ and 3′ splice sites. RNA(1–54), RNA(1–182), and RNA(1–236) were transcribed from plasmid pPIP85.A, which had been previously cleaved with either Scr F I, Dde I or Hind III, respectively, by the use of T7 RNA polymerase and a G(5′)ppp(5′)G dinucleotide primer (Pharmacia). DNA templates for transcription of RNA(55–236) and RNA(183–236) were generated by the polymerase chain reaction (PCR) with primers that inserted T7 promoters immediately 5′ to nucleotides 56 and 184, respectively, and M13 reverse primer. After Hind III cleavage of the PCR products, RNA(55–236) and RNA(183–236) were transcribed from the appropriate templates with GpG dinucleotide primers (*17*), which resulted in incorporation of an extra guanosine at the 5′ terminus of each RNA, thereby regenerating the entire intact sequence.
17. The dinucleotide primers, rG_prG, dG_prG, and mG_prG, were synthesized on a solid phase support from commercially available reagents (Applied Biosystems; Milligen; Glen Research). After cleavage and deprotection, they were purified by C_{18} reversed-phase HPLC. Maximal primer incorporation by T7 RNA polymerase was obtained at a ratio of primer to GTP of 5:1 (500 μM GpG and 100 μM GTP) in the transcription reactions.
18. T1 ribonuclease requires a free 2′-OH to form the 2′,3′-cyclic phosphate cleavage product; see K. Takahashi and S. Moore, in *The Enzymes*, P. Boyer, Ed. (Academic Press, New York, 1982), vol. 15, pp. 435–468.
19. Splicing reactions were performed at 30°C in 20 to 40 percent HeLa nuclear extract, 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and RNasin (Promega) at 1 U/μl [P. J. Grabowski, R. A. Padgett, P. A. Sharp, *Cell* **37**, 415 (1984)]. Nuclear extracts were prepared as described [J. D. Dignam, R. M. Lebowitz, R. G. Roeder, *Nucleic Acids Res.* **11**, 1475 (1983)]. After the incubations, RNA's were extracted and subjected to denaturing polyacrylamide gel electrophoresis.
20. A possible explanation for the electrophoretic mobility anomalies could lie in the tris-borate (pH 8.3) buffer system used for electrophoresis. Chelation of boric acid by the terminal cis-diol of the all-ribo E1 RNA might increase its bulk, and therefore decrease its mobility, relative to the 2′-H and 2′-OCH_3 E1 RNA's.
21. A similar effect had been observed for nucleotide mutations at the 5′ splice site of rabbit β-globin large intron which permitted the first step of splicing, but inhibited the second step [M. Aebi, H. Hornig, R. A. Padgett, J. Reiser, C. Weissman, *Cell* **47**, 555 (1986)]. In that study, an unexplained species with electrophoretic mobility similar to that of excised lariat IVS was observed without concurrent accumulation of a ligated exon product.
22. Hydrolysis would result from the increased usage of OH^-, rather than the free 3′-OH of E1, as the nucleophile at the 3′ splice site. Such hydrolysis can occur at the splice site phosphates of group I and group II introns [(*4*); R. van der Veen, J. H. J. M. Kwakman, L. A. Grivell, *EMBO J.* **6**, 3827 (1987)].
23. J. C. S. Noble, C. Prives, J. L. Manley, *Nucleic Acids Res.* **14**, 1219 (1986); X. Wang and R. A. Padgett, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7795 (1989).
24. D. A. Usher, D. I. Richardson, Jr., F. Eckstein, *Nature* **228**, 663 (1970); F. M. Richards and H. W. Wyckoff, in *The Enzymes*, P. Boyer, Ed. (Academic Press, New York, 1971), vol. 4, pp. 647–807; P. Blackburn and S. Moore, *ibid.* (1982), vol. 15, pp. 317–433.
25. J.-P. Perreault, T. Wu, B. Cousineau, K. K. Ogilvie, R. Cedergren, *Nature* **344**, 565 (1990) and references.
26. D. Herschlag and T. R. Cech, *ibid.*, p. 405.
27. See for example, G. J. Quigley and A. Rich, *Science* **194**, 796 (1976); A. C. Dock-Bregeon *et al.*, *Nature* **335**, 375 (1988).
28. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
29. G. F. Joyce, *Nature* **338**, 217 (1989); H.-Y. Mei, T. W. Kaaret, T. C. Bruice, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9727 (1989).
30. T. Inoue, F. X. Sullivan, T. R. Cech, *J. Mol. Biol.* **189**, 143 (1986); P. S. Kay and T. Inoue, *Nature* **327**, 343 (1987); M. D. Been and A. T. Perrotta, *Science* **252**, 434 (1991).
31. B. C. Rymond and M. Rosbash, *Nature* **317**, 735 (1985).
32. C. W. J. Smith, E. B. Porro, J. G. Patton, B. Nadal-Ginard, *ibid.* **342**, 243 (1989).
33. Base substitutions at the conserved 3′ splice site guanosine can significantly inhibit the formation of lariat intermediate for some mammalian introns [see ref. in (*21*)]. It is possible that recognition of the guanine base at the 3′ splice site is more important for the first step of splicing than recognition of its ribose moiety.
34. B. L. Bass and T. R. Cech, *Nature* **308**, 820 (1984); F. Michel, M. Hanna, R. Green, D. Bartel, J. Szostak, *ibid.* **342**, 391 (1989).
35. B. L. Bass and T. R. Cech, *Biochemistry* **25**, 4473 (1986).
36. For each band in every lane, an individual background value was determined from the area in the same lane immediately above that band. Therefore, any increase in background in lower portions of the gel due to nucleolytic degradation was fully compensated for in the rate calculations.
37. M. M. Konarska, P. J. Grabowski, R. A. Padgett, P. A. Sharp, *Nature* **313**, 552 (1985).
38. In this experiment, the appearance of the E1 and IVS-E2 RNA intermediates was somewhat delayed (by approximately 10 minutes) in the 5′ splice site-(2′-OCH_3) time course compared to that of the all-ribo. After this delay, however, the rates of accumulation of the 5′ splice site-(2′-OCH_3) E1 and IVS-E2 RNA's were the same as the all-ribo rates (Table1).
39. Except for lanes 9 and 14, control RNA's were generated by digestion of singly labeled E1-IVS-E2 RNA's, containing the appropriate modification at either the 5′ or 3′ splice site, with RNase A, T1 ribonuclease or calf intestinal phosphatase (CIP) or a combination of these. All-ribo ApGpG (lane 9) was generated by joining 5′-^{32}P-labeled $_prG_prG$ to RNA(1–182) with T4 RNA ligase, followed by digestion with RNase A and CIP. ApG containing the 2′-H modification on guanosine (lane 14) was made by extending RNA(1–182) by one nucleotide with Klenow DNA polymerase and [α-^{32}P]dGTP on a cDNA template that was complementary to the 3′ end of the RNA, followed by digestion with RNase A. No control RNA could be easily made for ApG containing the 2′-OCH_3 modification on the guanosine. All digestions were performed in 10 mM tris-HCl (pH 7.5), 1 mM EDTA and *E. coli* tRNA (0.7 μg/μl) for 20 minutes at 37°C. Enzyme concentrations were: 0.1 mg/ml, RNase A (BMB); 10 U/μl, T1 ribonuclease (BMB); or 0.1 U/μl, CIP (BMB).

40. We thank R. Cook, H. LeBlanc, and S. Schultz of the MIT Biopolymers Facility for synthesizing and purifying the GpG dinucleotides; D. Black, U. L. RajBhandary, and J. Stubbe for helpful discussions; and members of the Sharp laboratory (J. Crispino, T. Kristie, R. Meyers, and C. Query) for critical reading of the manuscript. Supported by a Helen Hay Whitney postdoctoral fellowship (M.J.M.), NIH grant RO1-GM34277 (P.A.S.), and NCI core grant P30-CA14051 (Center for Cancer Research).

27 January 1992; accepted 15 April 1992