Splicing of Messenger RNA Precursors

PHILLIP A. SHARP

A general mechanism for the splicing of nuclear messenger RNA precursors in eukaryotic cells has been widely accepted. This mechanism, which generates lariat RNAs possessing a branch site, seems related to the RNA-catalyzed reactions of self-splicing introns. The splicing of nuclear messenger RNA precursors involves the formation of a multicomponent complex, the spliceosome. This splicing body contains at least three different small nuclear ribonucleoprotein particles (snRNPs), U2, U5, and U4+U6. A complex containing precursor RNA and the U2 snRNP particle is a likely intermediate in the formation of the spliceosome.

THE DISCOVERY OF RNA SPLICING AND INTRONS IN 1977 heralded a new era in the study of the molecular biology of eukaryotic gene expression. The complexity of gene organization, the combinatorial possibilities of assembling different coding exons from an RNA precursor, and the novelty of the RNA-splicing process clearly indicated that eukaryotic molecular biology would be fascinatingly different from that of prokaryotic systems. Since 1977, a wealth of novel results have emerged from the area of eukaryotic gene structure and expression. Here, we will focus on recent results concerning the splicing of messenger RNA (mRNA) precursors (*1*).

Questions

Several global questions are central to the study of RNA splicing. The first of these is the basis of the specificity of excision of intervening sequences. This can be divided into two subquestions: What local information specifies the particular phosphodiester bonds to be cleaved and ligated, and what additional information specifies the excision of introns that are thousands of base pairs in length? It is now clear that small nuclear ribonucleoprotein particles (snRNPs) are crucial factors for sequence recognition during splicing. Changes in such factors could regulate splicing patterns observed in different cell types or physiological states (*2*). Another important question is the chemistry of the splicing reactions: What are the mechanisms of the endonucleolytic cleavage and ligation steps, and are these reactions catalyzed by RNA or protein components? A tentative relation between self-splicing introns, which are excised by RNA catalysis (*3*), and the splicing of mRNA precursors has implications for the origin and antiquity of introns. Finally, what is the relation between RNA splicing and transport of RNA from the nucleus? Answers to these questions are not yet known but recent results suggest some interesting possibilities.

The author is director of the Center for Cancer Research and professor in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

Intermediate in the Splicing Reaction

A single general mechanism is thought to be responsible for the splicing of all nuclear mRNA precursors in mammalian, plant, and yeast cells. Analysis of the splicing of radioactive substrate RNA in reactions containing nuclear extracts of either mammalian (*4–7*) or yeast cells (*8*) revealed the generation of lariat RNAs. A lariat RNA contains a site where the molecule branches (*9*). The excised intervening sequences (IVS) are released as a lariat RNA with the terminal guanosine residue linked through a 2′–5′ phosphodiester bond to an adenosine residue within the intron (Fig. 1). The branch site is typically 20 to 50 nucleotides upstream of the 3′ splice site. A kinetic intermediate forms during splicing (*4, 6*). This intermediate consists of two RNAs, the 5′ exon (E1) and a lariat form of the IVS-3′ exon (E2) (Fig. 1). Cleavage at the 5′ splice site and formation of the branch is the first covalent modification of the precursor during splicing. The cleavage and branch formation reactions have not been resolved kinetically and are thought to occur simultaneously. Similarly, the cleavage reaction at the 3′ splice site is thought to occur simultaneously with ligation of the two exons. The phosphate moieties at both the 5′ and the 3′ splice sites are conserved in the products (*5–7, 10*). Thus, both steps may be trans-esterification reactions, where a hydroxyl group reacts with a phosphodiester bond displacing a hydroxyl group while forming a new phosphodiester bond.

Branched RNAs and Self-Splicing Introns

The introns of yeast mitochondria can be assigned to two groups on the basis of two different sets of common consensus sequences (*11*). The common consensus sequences of the group I type are also found in the intron of the nuclear gene for ribosomal RNA in the protozoan *Tetrahymena*. As Cech *et al.* discovered, RNA precursors containing this intron self-splice in the absence of protein (*12*). It is now assumed that all group I introns are processed by a similar RNA-catalyzed reaction the rate of which is enhanced by the binding of proteins. Until recently, the nature of the splicing reactions for the group II introns remained an enigma. This has changed with the discovery that group II introns also have self-splicing activity (*13, 14*). However, in contrast to the group I type self-splicing reaction, the group II self-splicing process produces branched RNAs.

A schematic comparison of the self-splicing reactions of group I and II introns is shown in Fig. 2. Group I introns use a guanosine cofactor to self-splice in a two-step pathway (*3*). Both reactions in the pathway are trans-esterifications and thus the phosphate moieties at the splice sites are retained. The self-splicing of group II introns occurs in the absence of a nucleotide cofactor since the 2′ hydroxyl group at the branch site participates in the first trans-esterification reaction (*13*). This produces the lariat RNA that is subsequently found in the excised intron. Shown at the right in Fig.

2 is a similar schematic for the splicing process of nuclear mRNA precursors. A comparison of all three processes shows striking parallels. Each involves a two-step pathway; the first step is cleavage at the 5′ splice site. In all three processes, the phosphate moieties at the splicing sites are conserved in the products. Both the group II type and the nuclear mRNA precursor processes form branches at a site within the intron. These similarities strongly suggest that the splicing process of nuclear mRNA precursors is closely related to that of RNA-catalyzed self-splicing reactions (*15, 16*). This relation could reflect a common evolutionary origin so that the mRNA precursor process might be descended from the putatively more primitive RNA-catalyzed process.

As mentioned above, self-splicing introns can be assigned to two groups on the basis of common consensus sequences. The total number of nucleotides in the conserved consensus sequences in the group I and II introns is approximately 30 to 50 nucleotides, respectively (*11*). Secondary structures in the vicinity of these consensus sequences are also conserved although the primary sequences within these structures are not. In total, both consensus sequences and conserved secondary structures can be accommodated in 150 nucleotides of RNA. Mutational studies have shown that both the consensus sequences and secondary structures identified in group I introns are important for self-splicing (*3*). This work strongly suggests that a short core of conserved RNA sequences forms a catalytic pocket where the sequences at the splice sites and the guanosine cofactor bind and react. Although the catalytic activity of this core structure was first demonstrated by an intramolecular reaction, self-splicing, the core RNA also has activity in intermolecular reaction. For example, it has recently been shown that the core structure of a group I intron can catalyze cleavage at the 5′ splice site of a substrate RNA (*17*). Thus, covalent linkage of the catalytic core RNA with substrate RNA that contains splice sites is almost certainly unnecessary for RNA-catalyzed splicing.

The process of splicing of nuclear mRNA precursors seems evolutionarily related to the self-splicing process, and one or both of the two steps in the nuclear process may prove to be RNA catalyzed. Any catalytic RNA involved in the splicing of mRNA precursors could not be part of the precursor RNA, since mutational analysis has shown that the only intron sequences essential for splicing are the limited consensus sequences near the 5′ and 3′ splice sites (*18*). The obvious candidates for specifying this hypothetical catalytic RNA structure are the highly evolutionary conserved small nuclear RNAs (snRNAs). These RNAs are present in the nucleus as ribonucleoprotein particles (*19, 20*). Several of these snRNP particles have recently been shown to be essential for splicing of mRNA precursors in vitro.

snRNPs and Splicing

The nucleoplasm of animal cells contains five types of snRNP particles with an abundance greater than 10^5 per cell and probably a number of less abundant types (*20*). These particles possess between five and nine polypeptides in addition to the RNA component, snRNA. U1, U2, and U5 snRNPs appear to contain a single RNA per particle, whereas the U4+U6 snRNAs can coexist within a single particle (*21*). The total complement of snRNAs within animal cells is not known since many low abundance species may have escaped detection. As an example, another snRNA, U7, was discovered only through a biochemical assay for the processing of the 3′ terminus of a histone mRNA precursor (*22*). This species is present in sea urchin nuclei at approximately 5×10^4 copies per cell. The snRNAs share some distinctive properties, a trimethyl cap structure at the 5′ terminus of the RNA moiety, a common set of core

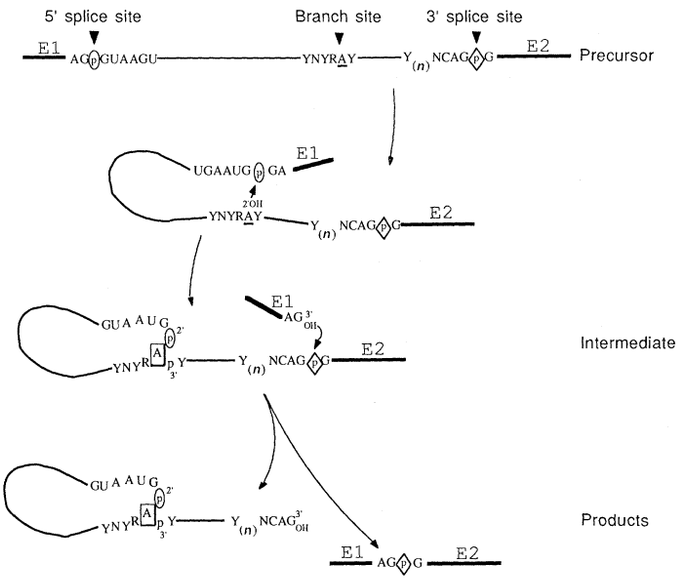

Fig. 1. Splicing mechanism of the mRNA precursor. A prototype precursor RNA is drawn with the intervening sequences or intron sequences spanning from the indicated 5′ and 3′ splice sites. The intervening sequences are flanked by the 5′ exon (E1) and 3′ exon (E2). Consensus sequences are indicated at the splice sites and branch site (Y, pyrimidine; R, purine; and N, any base). The fate of the phosphate moieties at the 5′ and 3′ splice sites during the reaction can be deduced from following the circled p and p surrounded by a diamond, respectively. The two RNAs of the intermediate are diagrammed in the third line. At the bottom of the figure, the two products of the reaction, the lariat form of the excised intervening sequences and the spliced exons, are shown.

polypeptides recognized by poly- or monoclonal antibodies of the Sm type (*23*), and an internal RNA sequence of the type AUUUUUG. This sequence is thought to be responsible for the binding of some of the core polypeptides and, subsequently, further methylation of the cap (*24*). From these criteria, other less abundant snRNAs have also been identified in mammalian cells (*25*).

The first connection between splicing and snRNA was the hypothesis of Lerner *et al.* (*26*) and Rogers and Wall (*27*) that the 5′ terminal region of U1 snRNA recognizes the 5′ splice site by sequence complementarity. Until recently, biochemical evidence in support of this hypothesis was only strongly suggestive. Conclusive evidence has now been provided by the demonstration that the splicing efficiency of mutant 5′ splice sites can be increased in vivo by expression of mutant U1 snRNA genes with compensatory mutations in their 5′ sequences (*28*). Thus, it can now be accepted that U1 snRNP recognizes sequences at the 5′ splice site and promotes the splicing reaction (Fig. 3).

The second most abundant particle in mammalian nuclei is the U2 snRNP. When substrate RNA is incubated in cell extracts, a complex containing this particle forms on sequences upstream of the 3′ splice site that encompass the branch site and at least part of the polypyrimidine tract (*29, 30*; see Figs. 1 and 3). Formation of the U2 snRNP complex requires addition of adenosine 5′-triphosphate (ATP) to the reaction (*29, 31*). The nature of the sequence specificity for binding by U2 snRNP is not yet clear. Mutations in the highly conserved *AG* sequences at the 3′ splice site will prevent formation of the U2 snRNP complex on a substrate RNA that lacks a 5′ splice site sequence (*32*). However, a U2 snRNP complex will form on RNA from such a mutant if the substrate RNA contains a wild-type 5′ splice site sequence. These observations suggest that (i) recognition of the 5′ splice site, almost certainly by U1 snRNP, facilitates or stabilizes binding by U2 snRNP, and (ii) in the absence of recognition of the 5′ splice site, formation of the U2 snRNP

complex depends critically on sequences at the 3′ splice site. The binding of U2 snRNP is probably not critically dependent on specific sequences flanking the branch site, since these sequences vary radically among functional introns (*33, 34*). Deletions in the polypyrimidine tract inactivate both binding of U2 snRNP (*35, 36*) and splicing (*33, 34*). Thus the polypyrimidine tract as well as the *AG* dinucleotide at the 3′ splice site might be important for interactions with U2 snRNP.

It is apparent that U2 snRNA does not recognize the polypyrimidine tract and 3′ splice site by extensive sequence complementarity (see below). In this regard, it is interesting that a protein has recently been identified that has binding specificity for sequences in or near the polypyrimidine tract at the 3′ splice site (*37*). Perhaps this protein is part of the complex formed with substrate RNA and U2 snRNP. It has been suggested that U5 snRNP can bind the 3′ splice site in an ATP-independent fashion (*31*). If this is the case, then perhaps U5 snRNP binding to the 3′ splice site could also facilitate the ATP-dependent formation of the U2 snRNP complex.

The substrate RNA–U2 snRNP complex is remarkably stable, surviving days of storage at 4°C. This stability suggests that formation of the U2 snRNP complex on a nuclear precursor RNA in vivo could commit that region of the RNA to the intron role in a splicing reaction. Formation of these types of complexes along a precursor RNA would then specify the number of introns to be excised. The temporal order of formation of U2 snRNP complexes on precursors containing multiple introns is unknown. Two extreme possibilities are that (i) shortly after transcription, U2 snRNP complexes form on all introns, and the order of splicing of the multiple introns is specified by subsequent slower events or (ii) a U2 snRNP complex forms on any particular intron immediately before splicing, and the rate of formation of this complex dictates the order of excision of the multiple introns. Resolution of this question seems of central importance for understanding the regulation of splicing.

Spliceosome

Splicing only occurs after formation of a splicing body, or spliceosome (*38*). The experimental definition of a spliceosome is a multicomponent complex that forms on the precursor RNA before the first cleavage step. A strong indication that such a multicomponent complex might exist was the bipartite RNA structure of the intermediate in splicing (Fig. 1). Such a complex would hold the two RNAs juxtaposed for the second step in splicing. When tested experimentally, the two RNAs of the intermediate cosediment as part of a complex, the spliceosome, at 40*S* to 60*S*. Treatment of the complex by heparin, a polyanion, reduces the sedimentation rate of the mammalian spliceosomes to 35*S*, probably by removing proteins or components that are electrostatically bound (*39*). Electrophoresis of splicing reactions on native polyacrylamide gels has also been used to resolve spliceosome-type complexes. In some cases, the complex is treated with heparin before electrophoresis (*30*). Under these conditions, the U2 snRNP complex described above can be readily resolved from a slower migrating spliceosome complex. Kinetically, formation of the U2 snRNP complex on the 3′ splice site precedes formation of the spliceosome containing the intermediate RNA forms.

The spliceosome has also been purified by affinity chromatography of biotin conjugates of substrate RNA (*39*). Incubation of this RNA in a splicing reaction results in formation of a spliceosome that can be purified by the combination of centrifugation through glycerol gradients and affinity chromatography on streptavidin-Sepharose. Purification by this method of a heparin-treated spliceosome clearly showed that the spliceosome contained, along with substrate RNA, approximately equal molar amounts of the snRNAs U2, U4, U5, and U6 (*39*). Surprisingly, the spliceosome did not contain significant levels of U1 snRNA, although U1 snRNP function appears essential for spliceosome formation. This observation would seem to conflict with recent evidence showing that antigens associated with the U1 snRNP are found in complexes containing lariat RNAs (*36*). It is potentially possible that these antigens may be present in the spliceosome in the absence of U1 snRNA. Alternatively, U1 snRNP might have been lost from the spliceosome during purification by affinity chromatography on gel electrophoresis. If U1 snRNP is not present, then this particle would play the role of a transient factor in the formation of the spliceosome.

The combination of sedimentation and affinity chromatography has also been used to purify the substrate RNA–U2 snRNP complex. The only snRNA found in this complex is U2 (*39*). Again, it is somewhat surprising that this initial complex encompassing the 3′ splice site does not also contain U5 snRNA, which has been proposed as the initial recognition factor for the 3′ splice site.

The copurification of the U4 and U6 snRNAs with the spliceosome suggests that the particle containing these snRNAs is important in splicing. An essential role for the U4+U6 snRNP in splicing has also been suggested from the results of inactivation experiments. Degradation of U4 snRNA by digestion of nuclear extracts with a combination of ribonuclease H and an oligodeoxynucleotide complementary to sequences of U4 snRNA inactivates the splicing capacity of the extract (*40*). Similar experiments had shown that U1 snRNA and U2 snRNA are important for splicing in vitro (*29, 41*). As yet, it has not been possible to use specific probes to inactivate U5 snRNA. These types of experiments seldom inactivate splicing capacity completely. This may be explained by the presence of some residual intact snRNA that escapes cleavage or by only a partial requirement for a given snRNP particle. Thus, it is difficult to distinguish with this protocol between an absolute requirement for a snRNP in splicing and a situation where a particular snRNP simply increases the rate of the reaction.

The nature of the specificity for the binding of U4+U6 and U5 snRNPs during formation of the spliceosome structure is unclear. Incubation of nuclear extracts in the absence of exogenous substrate RNA results in the accumulation of a multi-snRNP complex of U4+U6 and U5 snRNPs (*42*). Whether the binding of this multi-snRNP complex is primarily to the U2 snRNP or to sequences in the precursor RNA or to both remains to be determined. If U1 snRNP is not part of the final spliceosome, then some other component in the spliceosome must also recognize sequences at the 5′ splice site.

The presence of multiple snRNPs in the spliceosome suggests that snRNP-snRNP interactions may dictate the backbone of this structure. The rules for these snRNP-snRNP interactions will probably soon emerge from resolution of multi-snRNP complexes in extracts by gel electrophoresis (*30*). The multi-snRNP nature of the spliceosome also raises the possibility that different introns may be recognized by different combinations of snRNPs. This provides a mechanism by which the splicing of different introns could be individually regulated, that is, by modulation of the activity of a snRNP required for the splicing of only a subset of introns.

snRNAs in Yeast Spliceosome

Introns in the yeast *Saccharomyces cerevisiae* are typically short and contain highly conserved 7- and 8-bp sequences at their 5′ splice site

and branch site, respectively, and a simple *AG* at their 3′ splice site (*43*). These sequences are sufficiently related to the consensus sequences of introns in metazoa to indicate that the splicing processes in the two systems are closely related. Several small nuclear RNAs with the characteristic trimethyl cap of snRNAs have been identified in *S. cerevisiae* (*44*). At least three of these copurify with spliceosome complexes (*45*). As yet, only one of these snRNAs has been sequenced, and the results are remarkable. The yeast snRNA is 1175 nucleotides long and contains extensive homology near its 5′ terminus to the mammalian U2 snRNA, 43 identical bases in 47 positions (*46*). This RNA is clearly the yeast analog of U2 snRNA. In addition to the homology to U2 snRNA, the yeast snRNA also contains much shorter patches of homology to mammalian U4, U6, and U5 snRNAs. Thus, this yeast snRNA might be the equivalent of a poly-snRNA encompassing the combination of snRNAs found in a mammalian spliceosome. It is somewhat surprising that the large yeast snRNA does not contain regions complementary to the highly conserved sequences at the 5′ splice site and branch site of yeast introns. Thus, either proteins or the other yeast snRNAs in the spliceosome recognize the conserved intron sequences.

All fungi may not specify the same assortment and structure of snRNAs. For example, the yeast *Schizosaccharomyces pombe*, which is only very distantly related to *Saccharomyces cerevisiae*, expresses a U2 snRNA that is of the same length and general structure as the vertebrate U2 snRNA (*46*). Thus, this organism may have a splicing apparatus more similar to that of mammals than that of *S. cerevisiae*. At least one intron from the SV40 small-T antigen transcript is accurately spliced in both mammalian cells and *Schizosaccharomyces pombe* (*47*).

Comparison of Splicing and Translation

The assembly of a spliceosome in either yeast or mammalian systems from stable ribonucleoprotein particles is reminiscent of the assembly of ribosomes for translation. In the latter case, a 40*S* ribosome subunit binds to a ternary complex of initiation factor (eif-2), initiation transfer RNA (met-tRNA), and guanosine 5′-triphosphate (GTP) (*48*). This complex then binds to a specific site on the mRNA, the initiation codon. Recognition of the cap at the 5′ end of the mRNA and ATP-dependent scanning are probably important in locating the initiation codon (*49*). Subsequently, the bound GTP is hydrolyzed to release guanosine 5′-diphosphate and eif-2 to allow the binding of the 60*S* ribosomal subunit. A number of initiation factors enhance the rate or are essential for these steps; most of these are transient factors, which are not constituents of the final 80*S* ribosome. For comparison, formation of the spliceosome requires ATP, is enhanced by recognition of the cap (*50*), involves the sequential binding of multiple ribonucleoprotein particles, is probably promoted by a number of transient factors that are not part of the final complex, and depends on recognition of limited tracts of sequences at the splice sites. These similarities suggest that some of the previous studies of translation may provide a useful paradigm for studies of splicing.

A major difference between the splicing process and translation is that the latter involves the assembly of only two ribonucleoprotein particles, whereas the former involves the assembly of at least four particles. This might reflect the fact that a spliceosome must recognize several sequence elements while a ribosome must only recognize one, the initiation codon. As mentioned before, this

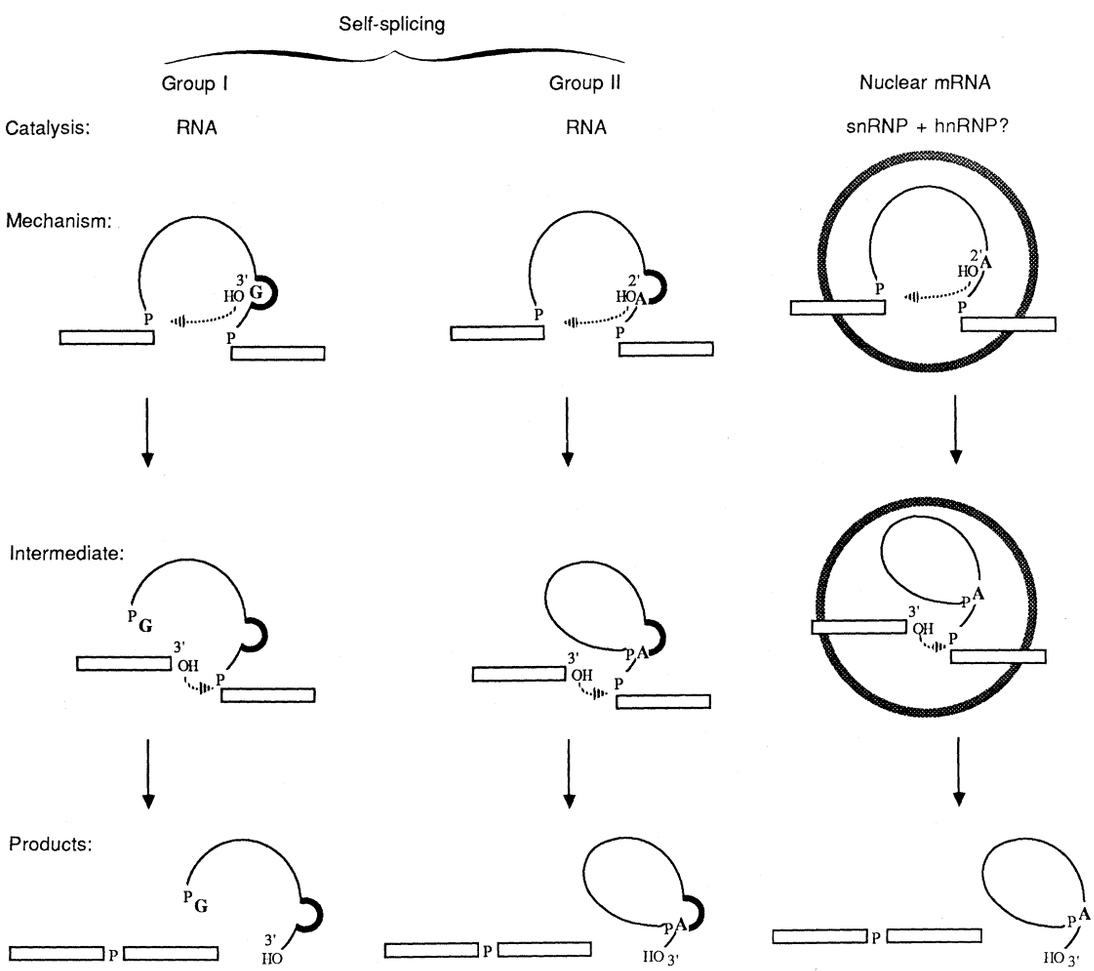

Fig. 2. Comparison of self-splicing and nuclear mRNA splicing mechanisms. The first column outlines the splicing mechanism of self-splicing introns of the group I type. This process is catalyzed by RNA structures within the intron (dark semicircle) utilizing a guanosine (G) cofactor. The second column outlines the splicing mechanism of self-splicing introns of the group II type. This process is also catalyzed by RNA structures within the intron (dark semicircle) but utilizing instead of a cofactor an adenosine residue (A) within the intron to form a lariat RNA. The third column outlines the mRNA precursor splicing mechanism diagrammed in Fig. 1. The large circle formed by the heavy line represents a multicomponent complex, the spliceosome, which promotes the splicing reaction. The fate of the phosphates at the 5′ and 3′ splice sites is indicated.

multi-snRNP aspect of splicing may be exploited in different cell states or introns to permit regulation of splicing of subgroups of introns by modulation of the activity of a particular snRNP. One may speculate that the splicing process might be evolutionarily older than the process of translation (*16, 51*). Processes such as splicing that can be catalyzed solely by RNA probably preceded those, such as translation, that utilize protein and, perhaps also, RNA catalysis. It may turn out that some components of the spliceosome and ribosome may be evolutionarily related. In this regard, it is somewhat surprising that prokaryotic organisms do not specify the mRNA splicing machinery. They have almost certainly shed this machinery during their evolution from ancestors that encoded it. Perhaps splicing by the mRNA process is inconsistent with coupled transcription-translation as occurs in prokaryotic organisms. It is interesting that other mechanisms of splicing, the self-splicing and the tRNA mechanisms, are compatible with prokaryotic systems.

Analysis in Vitro of Mutations in Splice Sites

The RNA sequences recognized during formation of the core spliceosome are those at the boundaries of the intron. At the 5′ splice site of mammalian introns this is the consensus sequence AG:GUAAGU. An extensive set of mutations in this sequence has been analyzed for activity in vivo and in vitro (*1*). Most changes decrease the rate of splicing and, in some cases, completely block synthesis of the mature RNA. It is somewhat surprising that mutations of either the *G* or *U* at the 5′ splice site allow the formation of the branch, thus the lariat RNA of the intermediate, but are defective for the second step in splicing (*34, 52*). Hence, the proximal *GU* sequence at the 5′ splice site is not necessary for precise cleavage. Arrest at the intermediate stage is also observed when the nucleotide at the branch site has been mutated from the conserved *A* residue to a *C* (*53*). These results clearly demonstrate that the splicing process involves two functionally separate steps. In addition, they suggest that the structure of the branch site is important for the second step, the reaction of the 5′ exon with the 3′ splice site (see Fig. 1). Perhaps a factor must recognize the branch site to promote processing at the 3′ splice site. Alternatively, the RNA structure at the branch site could be directly involved in the second reaction.

Mutations in the consensus sequences at the 3′ splice site can block, or reduce the rate of, both steps in splicing. As discussed above, formation of the U2 snRNP complex over the branch site and polypyrimidine tract is strongly but not completely dependent on recognition of the polypyrimidine tract and the *AG* at the 3′ splice site. This complex is essential for assembly of the complete spliceosome and thus for the first step in splicing. Mutations of the *AG* at the 3′ splice site can also result in accumulation of the intermediate. This is not so surprising, however, since the second step involves a reaction at the *AG* site.

Mutational studies have also shown that RNA secondary structures either involving splice sites or flanking splicing sites can affect the efficiency of splicing (*54*). In addition, sequence changes in adjacent exons also influence the rate of utilization of splice sites (*55*). These mutations in flanking sequences only modulate activity and thus may not be primary determinants for specificity. The importance of such sequences in specifying the processing of precursors in vivo is difficult to judge. Most of the splicing mutations isolated by phenotype selection alter the highly conserved sequences at the two splice sites and not sequences in flanking exons or introns (*1*).

For short introns of about 500 nucleotides, the sequence information in the consensus sequences at the two splice sites seems

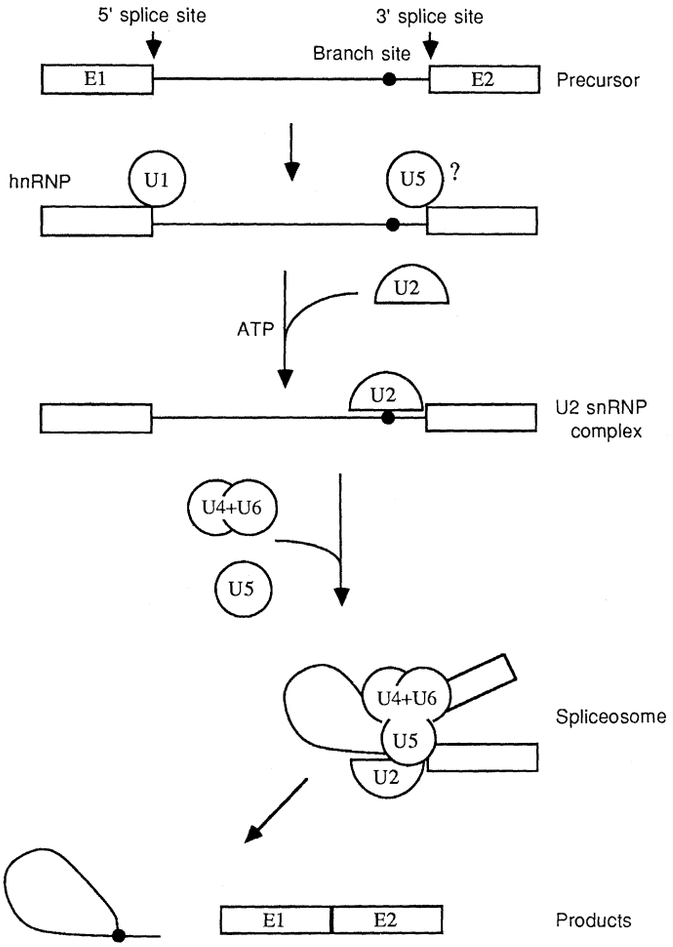

Fig. 3. Assembly of a spliceosome. The structure of the substrate RNA is shown on the first line. Addition of this RNA to a nuclear extract results in the recognition of the 5′ splice site by U1 snRNP. The precursor RNA is probably also associated with hnRNP core proteins. It has been suggested that U5 snRNP binds to the 3′ splice site at this stage. These interactions are indicated on the second line. Incubation of the substrate RNA in the presence of adenosine 5′-triphosphate (ATP) results in the rapid formation of a U2 snRNP complex as shown on the third line. Further incubation yields assembly of the spliceosome containing at least U2, U4, U5, and U6 snRNAs. The U4 and U6 snRNAs can be associated in one snRNP particle. The two RNAs characteristic of the intermediate in splicing, the 5′ exon and the lariat intervening sequence terminating in the 3′ exon RNAs, are found exclusively in the spliceosome.

sufficient to specify the reaction. However, for introns of 10,000 or more nucleotides, this limited specificity would seem to be inadequate. Introns of this length are not uncommon in mammalian genes and have also been observed in *Drosophila* and other organisms. The sequence specificity of the splicing of long introns has not really been studied either in vivo or in vitro; thus, one is left to speculate. Two types of additional sequence specificity for the splicing of long introns seem the most likely. It is possible to imagine that the precursor RNA is organized by the binding of proteins and snRNPs into a periodic structure where certain regions are brought into close proximity for processing. A defined structure has been determined for the RNP particles containing the long nuclear RNA precursors from the genes in the Balbiani rings (*56*). In addition, it is known that the majority of the heterogeneous nuclear RNA in cells is bound to an abundant group of proteins that limit nuclease sensitivity to sites at regular lengths of 500 to 700 nucleotides (*57*). One of these proteins, the heterogeneous nuclear RNP (hnRNP) *C* protein is important for splicing in vitro (*58*). The combination of periodic structure coupled with recognition of splice

sites might thus provide additional specificity for splicing. It should be noted, however, that specificity based on the generation of a unique structure extending over significant lengths of the precursor would place restrictions on the length of introns. This would suggest that deletion of sequences from the middle of long introns might affect their accuracy of splicing.

Another source for specificity in the splicing of long introns could be provided by additional sequences specifying the binding of snRNPs within the intron. These snRNPs could be of the abundant class such as U1, U2, U4+U6, and U5 snRNP or of less abundant and yet undefined classes. The presence within long introns of short tracts of sequences that are important for the specificity of splicing would probably have escaped detection in previous studies. Thus in one model for the processing of long introns a backbone formed by the interactions of snRNPs specifies the incorporation of the two correct splice sites into the spliceosome.

RNA Splicing and Nuclear Matrix

Interactions of snRNP with snRNP are critically important for splicing and might be of central importance to the structure and function of the nucleus. Many of these particles are present at over 10^5 to 10^6 units per nucleus. In addition, the total number of different snRNPs in mammalian cells has not been determined. As Riedel *et al.* (*44*) have recently pointed out, the yeast *Saccharomyces cerevisiae* contains at least 24 different snRNAs at an abundance of 10 to 500 copies per cell. Mammalian cells could potentially contain many times this number since low abundance snRNAs have not been analyzed. In addition to splicing, several potential roles for snRNP-snRNP interactions in nuclei are easily conjectured. In fact, evidence suggests that the polyadenylation reaction, cleavage of the precursor RNA and addition of a polyadenylate tract to create the 3′ end of the mRNA, requires the activity of snRNPs (*59*).

It may be that a major part of the intranuclear structure is composed of snRNPs. A form of nuclear matrix has been defined operationally as material remaining after extraction of isolated nuclei with high salt (2*M* NaCl), nonionic detergent (1 percent Triton), and digestion of DNA [deoxyribonuclease (DNase) I treatment] (*60, 61*). When viewed by electron microscopy, such samples retain the overall outline of nuclei, but are composed entirely of fibrous material (*61*). Other procedures for isolation of a nuclear matrix have also been described (*62*). Transcriptionally active genes are preferentially associated with either type of matrix, and pulse-labeled mRNA precursor–type sequences are matrix bound. More importantly, unspliced or partially spliced nuclear mRNA precursors have been shown to be preferentially associated with the nuclear matrix (*63*). Finally, Zeitlin *et al.* (*64*) have reported that matrix-associated precursors of rabbit β-globin can be chased from the matrix when splicing is promoted by the addition of extracts. These observations suggest that nuclear mRNA precursors are matrix associated, and thus the snRNPs associated with these precursors may be constituents of the matrix. In fact, some data directly suggesting the retention of snRNPs in matrix preparations have been described (*65*).

Transport of mRNA from the nucleus to the cytoplasm has been a poorly explored subject. This process is very difficult to study with the available in vivo or in vitro assay systems. It is possible that a snRNP matrix could be involved in such transport. RNAs are known to exit from the nucleus by way of nuclear pores, and thus transport of nuclear RNA must be directional, to and through the pores. Providing a structural explanation for polarity of transport to the pores would be a major step in the analysis of transport. If such polarity involved assembly of snRNPs, so much the better.

REFERENCES AND NOTES

1. See for review of field until fall of 1985, R. A. Padgett *et al.* [*Annu. Rev. Biochem.* **55**, 1119 (1986)] and M. R. Green [*Annu. Rev. Genet.* **20**, 671 (1986)].
2. S. E. Leff, M. G. Rosenfeld, R. M. Evans, *Annu. Rev. Biochem.* **55**, 1091 (1986).
3. T. R. Cech and B. L. Bass, *ibid.*, p 599.
4. P. J. Grabowski, R. A. Padgett, P. A. Sharp, *Cell* **37**, 415 (1984).
5. R. A. Padgett, M. M. Konarska, P. J. Grabowski, S. F. Hardy, P. A. Sharp, *Science* **225**, 898 (1984).
6. B. Ruskin, A. R. Krainer, T. Maniatis, M. R. Green, *Cell* **38**, 317 (1984).
7. M. M. Konarska, P. J. Grabowski, R. A. Padgett, P. A. Sharp, *Nature (London)* **313**, 552 (1985).
8. H. Domdey *et al.*, *Cell* **39**, 611 (1984).
9. J. C. Wallace and M. Edmonds, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 950 (1983).
10. R.-J. Lin, A. J. Newman, S.-C. Cheng, J. Abelson, *J. Biol. Chem.* **260**, 14780 (1985).
11. F. Michel, A. Jacquier, B. Dujon, *Biochimie* **64**, 867 (1982); R. W. Davies, R. B. Waring, T. A. Ray, T. A. Brown, C. Scazzocchio, *Nature (London)* **300**, 719 (1982); F. Michel and B. Dujon, *EMBO J.* **2**, 33 (1983).
12. T. R. Cech, A. J. Zaug, G. J. Grabowski, *Cell* **27**, 487 (1981).
13. C. L. Peebles *et al.*, *ibid.* **44**, 213 (1986); R. van der Veen *et al.*, *ibid.*, p. 225.
14. C. Schmelzer and R. J. Schweyen, *ibid.* **46**, 557 (1986).
15. T. R. Cech, *ibid.* **44**, 207 (1986).
16. P. A. Sharp, *ibid.* **42**, 397 (1985).
17. J. W. Szostak, *Nature (London)* **322**, 83 (1986).
18. B. Wieringa, E. Hofer, C. Weissmann, *Cell* **37**, 915 (1983).
19. M. R. Lerner and J. A. Steitz, *ibid.* **25**, 298 (1981).
20. H. Busch, R. Reddy, L. Rothblum, Y. C. Choi, *Annu. Rev. Biochem.* **51**, 617 (1982).
21. P. Bringmann *et al.*, *EMBO J.* **3**, 1357 (1984); C. Hashimoto and J. A. Steitz, *Nucleic Acids Res.* **12**, 3283 (1984).
22. K. Strub, G. Galli, M. Busslinger, M. L. Birnstiel, *EMBO J.* **3**, 2801 (1984).
23. M. R. Lerner and J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5495 (1979).
24. I. Mattaj and E. DeRobertis, *Cell* **40**, 111 (1985).
25. R. Reddy, D. Henning, H. Busch, *J. Biol. Chem.* **260**, 10930 (1985).
26. M. R. Lerner, J. A. Boyle, S. Mount, S. L. Wolin, J. A. Steitz, *Nature (London)* **283**, 220 (1980).
27. J. Rogers and R. Wall, *Proc. Natl. Acad. Sci. U.S.A.* 77, 1877 (1980).
28. Y. Zhuang and A. Weiner, *Cell* **46**, 827 (1986).
29. D. L. Black, B. Chabot, J. A. Steitz, *ibid.* **42**, 737 (1985).
30. M. M. Konarska and P. A. Sharp, *ibid.* **46**, 845 (1986).
31. B. Chabot, D. L. Black, D. M. LeMaster, J. A. Steitz, *Science* **230**, 1344 (1985).
32. A. I. Lamond, unpublished results.
33. B. Ruskin, J. M. Greene, M. R. Green, *Cell* **41**, 833 (1985).
34. R. Reed and T. Maniatis, *ibid.*, p. 95.
35. B. Ruskin and M. R. Green, *ibid.* **43**, 131 (1985).
36. B. Chabot and J. Steitz, *Mol. Cell. Biol.*, in press.
37. J. Tazi *et al.*, *Cell* **47**, 755 (1986); V. Gerke and J. Steitz, *ibid.*, p. 973.
38. E. Brody and J. Abelson, *Science* **228**, 963 (1985); P. J. Grabowski, S. R. Seiler, P. A. Sharp, *Cell* **42**, 345 (1985); D. Frendewey and W. Keller, *ibid.*, p. 355.
39. P. J. Grabowski and P. A. Sharp, *Science* **233**, 1294 (1986).
40. S. M. Berget and B. L. Robberson, *Cell* **46**, 691 (1986); D. L. Black and J. A. Steitz, *ibid.*, p. 697.
41. A. Kramer, W. Keller, B. Appel, R. Luhrman, *ibid.* **38**, 299 (1984); A. R. Krainer and T. Maniatis, *ibid.* **42**, 725 (1985).
42. M. M. Konarska, unpublished results.
43. J. L. Teem *et al.*, *Nucleic Acids Res.* **22**, 8295 (1984).
44. N. Riedel, J. A. Wise, H. Swerdlow, A. Mak, C. Guthrie, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8097 (1986).
45. C. W. Pikielny and M. Rosbash, *Cell* **45**, 869 (1986).
46. M. Ares, Jr., *ibid.* **47**, 49 (1986).
47. N. Kaufer, V. Simanis, P. Nurse, *Nature (London)* **318**, 78 (1985).
48. K. Moldave, *Annu. Rev. Biochem.* **54**, 1109 (1985).
49. M. Kozak, *Adv. Virus Res.* **31**, 229 (1986).
50. M. M. Konarska, R. A. Padgett, P. A. Sharp, *Cell* **38**, 731 (1984); I. Edery and N. Sonenberg, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7590 (1985).
51. T. R. Cech, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4360 (1986); J. E. Darnell and W. F. Doolittle, *ibid.*, p. 1271; W. Gilbert, *Nature (London)* **319**, 618 (1986).
52. U. Vijayraghan *et al.*, *EMBO J.* **5**, 1683 (1986).
53. C. Weissmann, unpublished results.
54. D. Solnick, *Cell* **43**, 667 (1985); M. M. Konarska, R. A. Padgett, P. A. Sharp, *ibid.* **42**, 165 (1985); D. Solnick, *ibid.*, p. 157.
55. R. Reed and T. Maniatis, *ibid.* **46**, 681 (1986).
56. U. Skoglund, K. Andersson, B. Strandberg, B. Daneholt, *Nature (London)* **319**, 560 (1986).
57. G. Dreyfuss, *Annu. Rev. Cell. Biol.* **2**, 457 (1986).
58. Y. D. Choi, P. J. Grabowski, P. A. Sharp, G. Dreyfuss, *Science* **231**, 1534 (1986).
59. C. L. Moore and P. A. Sharp, *Cell* **41**, 845 (1985); C. Hashimoto and J. A. Steitz, *ibid.* **45**, 581 (1986).
60. R. Berezney and D. S. Coffey, *J. Cell Biol.* **73**, 616 (1977).
61. T. E. Miller, C.-Y. Huang, A. O. Pogo, *ibid.* **76**, 675 (1978).
62. E. G. Fey, G. Krochmalnic, S. Penman, *ibid.* **102**, 1654 (1986).
63. E. M. Ciejek, J. L. Nordstrom, M.-J. Tsai, B. O'Malley, *Biochemistry* **21**, 4945 (1982).
64. S. Zeitlin, A. Parent, S. Silverstein, A. Efstratiadis, *Mol. Cell. Biol.*, in press.
65. T. E. Miller, C.-Y. Huang, A. O. Pogo, *J. Cell Biol.* **76**, 692 (1978).

66. I would like to acknowledge the help of M. Siafaca for preparation of the manuscript, M. M. Konarska for preparation of the figures, and M. M. Konarska, P. Grabowski, A. Lamond, and S. Seiler for reading the manuscript. I also thank M. Ares, J. Steitz, and C. Guthrie for providing manuscripts in press. This work was supported by NIH grant GM34277, National Cancer Institute (NCI) PO1-CA42063, and partially from NCI core grant CA14051 to P.A.S.