Acta Biochimica Polonica

Vol. 45 No. 4/1998

869–881

QUARTERLY

This paper is dedicated to Professor Maciej Wiewiórowski
Review

Recognition of the 5′ splice site by the spliceosome

Maria M. Konarska

The Rockefeller University, New York, U.S.A.

Key words: pre-mRNA splicing, small nuclear RNP, splice site, spliceosome

The splicing of nuclear pre-mRNAs is catalyzed by a large, multicomponent ribonucleoprotein complex termed the spliceosome. Elucidation of the molecular mechanism of splicing identified small nuclear RNAs (snRNAs) as important components of the spliceosome, which, by analogy to the self-splicing group II introns, are implicated in formation of the catalytic center. In particular, the 5′ splice site (5′SS) and the branch site, which represent the two substrates for the first step of splicing, are first recognized by U1 and U2 snRNPs, respectively. This initial recognition of splice sites is responsible for the global definition of exons and introns, and represents the primary target for regulation of splicing. Subsequently, pairing interaction between the 5′SS and U1 snRNA is disrupted and replaced by a new interaction of the 5′SS with U6 snRNA. The 5′SS signal contains an invariant GU dinucleotide present at the 5′ end of nearly all known introns, however, the mechanism by which the spliceosome recognizes this element is not known. We have identified and characterized a specific UV light-induced crosslink formed between the 5′SS RNA and hPrp8, a protein component of U5 snRNP in the spliceosome that is likely to reflect a specific recognition of the GU dinucleotide for splicing. Because recognition of the 5′SS must be linked to formation of the catalytic site, the identification of a specific and direct interaction between the 5′SS and Prp8 has significant implications for the role of this protein in the mechanism of mRNA splicing.

°Supported by NIH Grant GM49044.

✉The Rockefeller University, 1230 York Avenue, New York, NY 10021, U.S.A.; tel.: (212) 327 8432, fax: (212) 327 7147; e-mail: konarsk@rockvax.rockefeller.edu

Abbreviations: snRNA, small nuclear ribonucleic acid; snRNP, small nuclear ribonucleoprotein particle; pre-mRNA, messenger RNA precursor; 5′SS, 5′ splice site; IVS, intervening sequence; Pu, purine.

SPLICEOSOME ASSEMBLY

Splicing takes place in a multicomponent, dynamic complex, termed the spliceosome, assembled on the pre-mRNA by the ordered association of small nuclear ribonucleoprotein particles (snRNPs). Five snRNP particles, U1, U2, U4, U5, and U6 snRNPs are involved in formation of the spliceosome, which creates the active center responsible for catalysis (reviewed in [1-9]). Spliceosome assembly requires specific recognition of splice site signals; the 5′ splice site (5′SS) consensus sequence, including a conserved GU dinucleotide at the 5′ end of the intron, and the 3′SS region, consisting of the branch site, polypyrimidine tract, and a conserved AG dinucleotide at the 3′SS [10]. The following description will focus mainly on the mammalian system; however, our understanding of the splicing mechanisms is also largely based on analysis of the highly homologous yeast (*Saccharomyces cerevisiae*) system [7, 11]. The initial stages of splicing involve formation of the commitment complex composed of U1 snRNP as well as non-snRNP proteins (see Fig. 1). Subsequently, splicing complex A is formed that contains U2 snRNP bound to pre-mRNA sequences around the branch site. Complex A is then converted into complex B by association with U4/U5/U6 snRNP. This step also involves an apparent destabilization of U1 snRNP from the spliceosome. Subsequent rearrangement leading to formation of complex C is associated with release of U4 snRNP from the spliceosome and the concomitant first step of splicing: cleavage at the 5′SS and formation of a lariat IVS-3′ exon intermediate. Finally, the lariat intron and spliced mRNA exons are generated. The lariat intron is found in complex I, containing U2, U5, and U6 snRNPs.

The multi-step process of spliceosome assembly is probably more complex than strictly required to form the functional catalytic site capable of carrying out two transesterification reactions of pre-mRNA splicing. However, this baroque design of the system

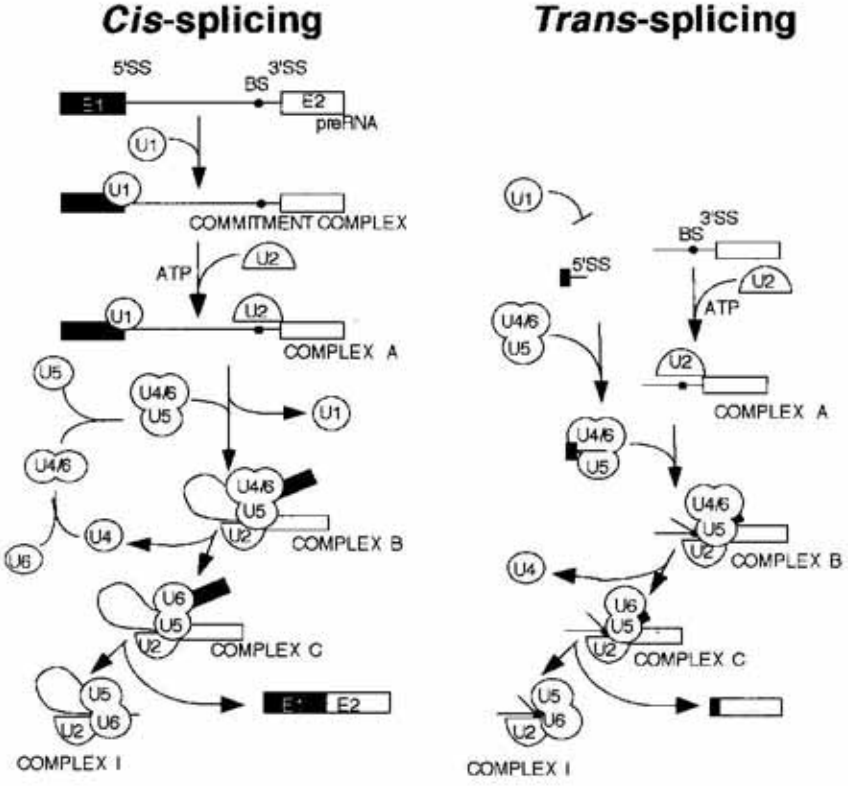

Figure 1. Schematic representation of the spliceosome assembly pathway.

Exons (E1 and E2) are represented by rectangles, introns by lines, and snRNP particles by circles. Relative positions of snRNPs in the splicing complexes are arbitrary.

allows for a highly precise and flexible control of splice site selection for the reaction. To maintain the reading frame of mRNA, introns must be removed with a single-nucleotide precision, yet they differ widely in size, number, and location within pre-mRNA, and they contain only short segments of conserved sequences. The specificity of recognition of splicing signals is provided by the combinatorial action of multiple factors that act at the individual splice sites and determine stability of the spliceosome. Thus, the complicated assembly process provides means of subtle and precise control of splice site selection and splicing. At least some of the spliceosome complexity serves regulatory purposes that provide the required specificity and selectivity of splicing.

ROLE OF snRNAs IN SPLICING

Multiple sequential interactions of snRNP particles and pre-mRNA are involved during spliceosome assembly (reviewed in [2–5, 7]). Briefly, the conserved 5′SS consensus sequence (G/GUPuAGU, where the / indicates the exon/intron border) is paired with the 5′ end of U1 snRNA within complex E or commitment complexes at the early stage of the process. In parallel, the branch site (5′UACUAAC3′) upstream of the 3′SS is recognized within complex A through base pairing with a segment of U2 snRNA (5′GUAGUA3′), producing a short duplex containing a branched adenosine, possibly bulged out from the helix. Subsequently, when U4/U5/U6 snRNP joins complex A to form complex B, U6 snRNA replaces U1 snRNA in the interaction with the intron at the 5′SS. At the stage of splicing catalysis, the extensive pairing within the U4/U6 snRNP is disrupted and replaced with new helices formed between U2 and U6 snRNAs. Thus, interaction between U2 snRNA paired to the branch site, with U6 snRNA paired to the intron sequence near the 5′SS, serves to bring together and properly position the two reactants for the first catalytic step (Fig. 2). The central loop of U5 snRNA interacts with exon sequences; however, in yeast this interaction is not essential for the first step of splicing [12, 13].

While the majority of pre-mRNA introns conform to the GU-AG consensus, a minor group of introns has recently been identified that exhibits significant differences in the se-

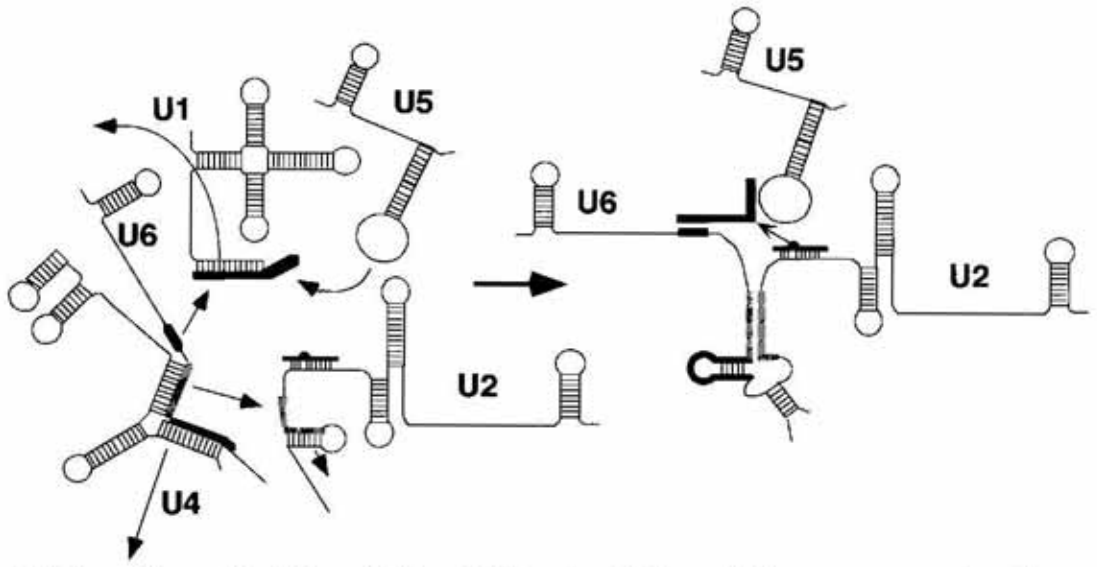

Figure 2. Schematic model of the snRNA:snRNA and snRNA:preRNA rearrangements at the catalytic center during the spliceosome assembly.

The 5′SS RNA oligonucleotide and the branch site substrates of the first step of splicing are depicted. While the scheme represents the U2-type snRNAs, equivalent rearrangements can be drawn for the U12-type pathway.

quences recognized at the exon/intron borders [14–16]. Importantly, splicing of these introns involves a distinct spliceosome composed of different snRNPs. Two of these novel snRNPs, U11 and U12, participate in the early recognition of the 5′ and 3′SS of this minor class (so called U12-type) of introns [14–16]. Interestingly, subsequent formation of the active spliceosome involves the same U5 snRNP as in the major class (U2-type) of spliceosomes, while U4 and U6 snRNPs are replaced with two minor class U4atac and U6atac snRNPs [17, 18]. While the primary sequences of U11, U12, U4atac and U6atac snRNAs are quite different from their U2-type counterparts, the most conserved sequences of U6 snRNA are present in U6atac snRNA [17]. Furthermore, the secondary structures of snRNAs and their interactions with other snRNAs and pre-mRNA are strikingly similar between the U2- and U12-type spliceosomes [16]. Thus, both spliceosomes are likely to follow the equivalent assembly pathways and conformational changes leading to catalysis.

Because of the mechanistic and structural similarities, splicing of pre-mRNA precursors and self-splicing group II introns are often thought to have originated from a common ancestor and proceed through RNA-based catalysis [19–21]. Biochemical and genetic data support the view that the catalytic mechanism of splicing depends on snRNA function and that it may have originated from the RNA world. According to this hypothesis, snRNAs represent descendants of functional domains of catalytic RNAs, which were capable of RNA splicing in the past [20]. At later stages of evolution, proteins began to interact with these RNA domains to accelerate the rate and increase the accuracy of splicing. With time, these catalytic domains began to act *in trans*, using associated proteins to facilitate assembly of larger, catalytically competent complexes. At present, proteins assist most, if not all, RNA-catalyzed reactions *in vivo*, although their precise role has not been determined in most cases. Thus, even if nuclear splicing has evolved from the RNA-catalyzed self-splicing process, it is formally possible that in addition to their regulatory functions in substrate recognition and building the structural framework of the complex, some spliceosomal proteins evolved even further to participate more directly in the catalytic process. In fact, the definition of the catalyst needs to be reevaluated in the context of an enzyme as complex as the spliceosome – composed of multiple proteins and RNAs that undergo a series of conformational changes in the process of splicing. Thus, before we can address the issue of the identity of the catalyst, we need to recognize other non-RNA important components of the spliceosome and learn more about the structure of its catalytic center.

SELECTION OF THE SPLICE SITES FOR SPLICING

One of the fundamental decisions in pre-mRNA processing is the selection of splice sites for the reaction. This global recognition of the exon:intron borders provides an early information concerning the subsequent precise definition of phosphodiester bonds that participate in the transesterification reactions of splicing. The early splicing complexes, involving the commitment complex and splicing complex A, select the 5′SS and the branch site regions, respectively. In complex A, the precise branch site is defined through base pairing interaction with a conserved sequence within U2 snRNA. On the other hand, the initial selection of the 5′SS occurs within commitment complexes and is primarily determined by the pairing between the 5′SS and the conserved sequence at the 5′ end of U1 snRNA.

Although U1 snRNP is required for spliceosome assembly, its association with the complex is apparently transient. One function of U1 snRNP involves the delivery of the 5′SS to the correct location in the complex, after

which U1 snRNP is destabilized. By analogy, it has been previously suggested that U4 snRNP functions by delivering U6 snRNP to the spliceosome and unblocking the catalytically inactive U6 snRNA through dissociation from the spliceosome at the transition from complex B to C [22]. Recent studies show that U1 snRNP is not strictly required for splicing of some introns and that SR proteins can promote efficient splicing of pre-mRNA in a U1 snRNP-independent fashion [23, 24]. These results can perhaps best be explained in light of the exon definition model [25, 26], which applies to most vertebrate organisms whose pre-mRNAs contain multiple short exons separated by long blocks of intervening sequences. According to this model, exons are first recognized, or defined, by binding of U1 and U2 snRNPs to the 5′ and 3′SS regions, respectively (Fig. 3). This association is stabilized by additional factors, including SR proteins, leading to the formation of commitment complexes [10, 24, 27–29]. Subsequently, active spliceosomes are formed upon further rearrangements that result in release of U1 snRNP and binding of U4/U5/U6 snRNP. In this model, interaction between complexes assembled at the two splice sites defines the sequence between them as the exon, while the remaining pre-mRNA sequences become by default defined as introns. Terminal exons require special mechanisms for their recognition. The 5′ terminal exons use the 7-methyl-guanosine cap structure at their 5′ termini as a recognition signal instead of the 3′SS signals typical of internal exons. The 3′ terminal exons relay on the recognition of polyadenylation signals in conjunction with the upstream 5′ SS elements. While the above model describes splicing of vertebrate pre-mRNAs that contain multiple short exons (average size – 137 nt) interspersed between long introns (even 100000 nt), it needs to be modified in the case of many lower eukaryotes that are characterized by long exons and few short introns. Thus, organisms with small introns, such as *S. cerevisiae*, use the intron definition, rather than the exon definition, as the initial mode of interaction between the splice sites. In this case, the process initiates with the interaction between complexes assembled across the intron (middle line in Fig. 3) and proceeds directly to formation of catalytically functional spliceosomes [26].

The exon definition model allows for accurate predictions of pre-mRNA splicing patterns in different organisms. For example, if

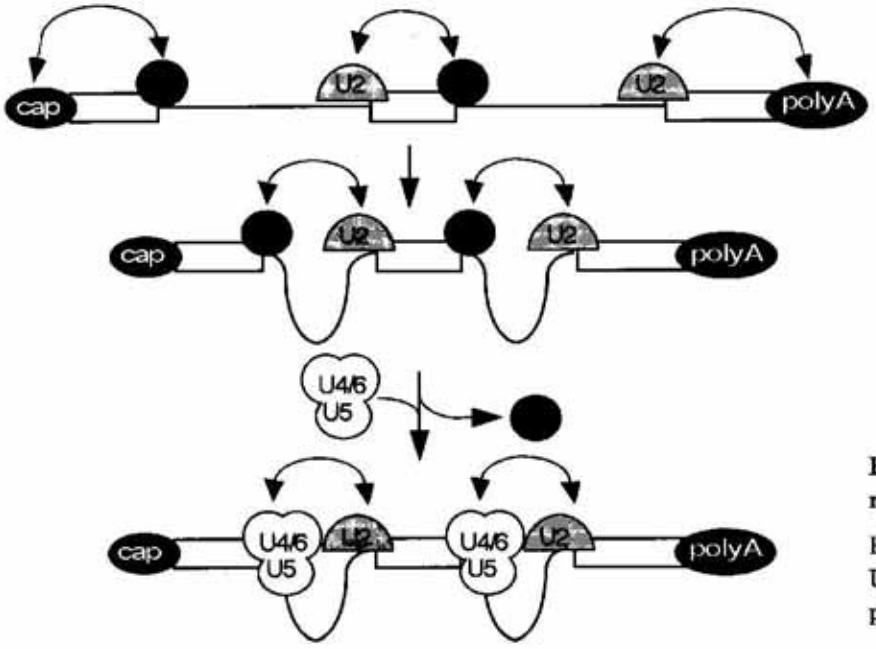

Figure 3. Exon definition model. See text for details.

Relative positions of U1, U2, and U4/U5/U6 snRNPs within complexes are arbitrary.

pre-mRNA undergoes splicing according to the intron definition mode, a mutation at the 5′ splice site flanking an intron will result in a loss of recognition of that intron and consequently retention of that intron in mature mRNA (Fig. 4A). In contrast, the analogous 5′SS mutation in pre-mRNA that undergoes splicing according to the exon definition mode will prevent recognition of the upstream exon, leading to exon skipping phenotype (Fig. 4B). Alternatively, a cryptic 5′SS may be activated in proximity of the mutated site, recreating a short, recognizable exon, and allowing for removal of the downstream intron. In fact, a survey of mammalian mutations available in the database indicates that exon skipping and activation of a cryptic splice site represent the most frequent phenotypes associated with splice site mutations, strongly supporting the exon definition model.

According to this hypothesis, U1 snRNP plays a regulatory role in the assembly process by directing proper selection of the 5′SS. In fact, most of the known cases of alternative splicing events are regulated at the stage of initial recognition of splice sites and formation of commitment complexes. Thus, while specific selection of splice sites is strictly necessary to maintain the proper composition of exons in the spliced mRNA, it is not directly required for the catalytic function of the spliceosome. As mentioned above, in some selected cases pre-mRNAs appear to undergo splicing in the U1 snRNP-independent fashion [23, 24, 30, 31].

A SIMPLIFIED *TRANS*-SPLICING SYSTEM

To study the recognition of the 5′SS at the later stages of spliceosome assembly, we have developed a simplified, functional *trans*-splicing system, which represents a variation on the well-characterized HeLa cell nuclear extract *in vitro* splicing system [32]. In this simplified system, a short RNA oligonucleotide containing the conserved 5′SS consensus sequence (5′SS RNA), together with a longer 3′SS RNA transcript, comprising the branch site, polypyrimidine tract, and the 3′ splice site, are used as substrates for splicing. Because of the complementarity between the

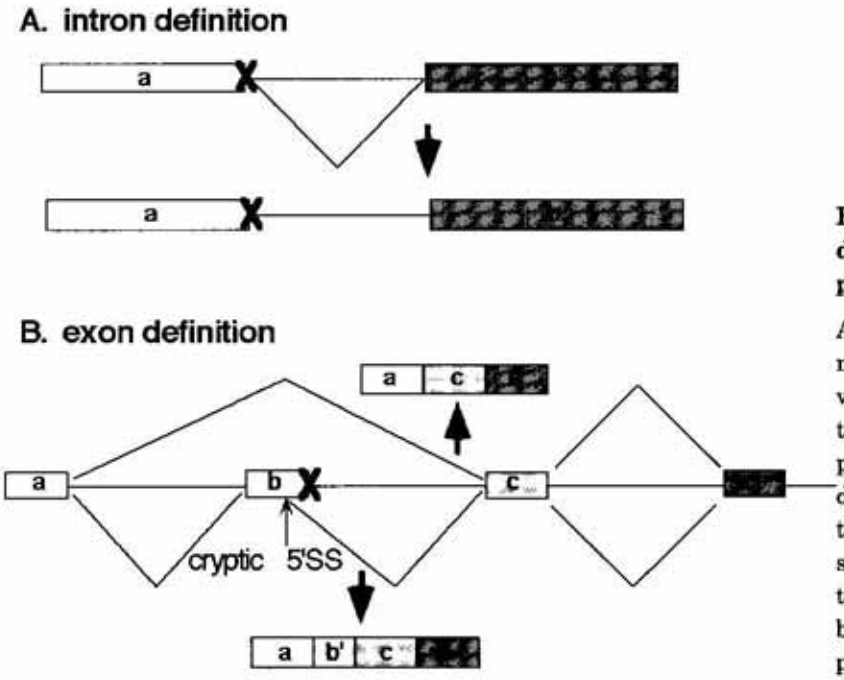

Figure 4. Selection of splice sites dictates the profiles of mRNA products.

A. According to the intron definition model, mutation at the 5′SS will prevent recognition of the intron, and thus will result in an intron retention phenotype. B. According to the exon definition model, a 5′SS mutation at the 3′ end of exon b will result in exon skipping, generating mRNA acd. Alternatively, a cryptic 5′ splice site may be activated, producing ab′cd mRNA product.

5′SS RNA and the 5′ end of U1 snRNA sequence, the 5′SS RNA is capable of binding to U1 snRNP [33]. Remarkably, in the absence of the 3′SS RNA, the 5′SS RNA can induce a specific association of U2 and U4/U5/U6 snRNPs, forming U2/U4/U5/U6 snRNP complex similar to splicing complex B [34]. Similarly, under *trans*-splicing conditions (i.e. in the presence of the 3′SS RNA), addition of the 5′SS RNA results in formation of a transient 5′SS RNA:U4/U5/U6 snRNP complex, which then stably associates with complex A to yield splicing complex B [32]. Thus, the 5′SS consensus sequence can be specifically recognized by U4/U5/U6 snRNP in the spliceosome, independently of U1 snRNP.

Because the stable 5′SS RNA:U1 snRNA interaction competes with formation of splicing complex B, disruption of the initial base pairing between the 5′SS RNA and U1 snRNA must be required for spliceosome assembly. Thus, to conveniently prevent binding of the 5′SS RNA to U1 snRNP and promote spliceosome assembly, a 5′SS DNA oligonucleotide competitor is included in *trans*-splicing reactions [32]. In the *trans*-splicing assay the initial recognition of the 5′SS by U1 snRNP is circumvented, resulting in U1-independent spliceosome formation and splicing [32, 33]. As a consequence, the specificity of exon definition provided by U1 snRNP is lost, and any two 5′ and 3′ splice site RNAs are joined in proportion to the relative affinity of their splice site signals and not their relative position within the precursor. Similarly, *trans*-splicing in Trypanosomes and nematodes does not seem to involve U1 snRNP particles [35]. In this system the spliced leader exon, SL RNA, may functionally substitute for the U1 snRNP particle, acting as a specialized snRNP particle itself [36].

In all splicing reactions, U1-dependent or not, the 5′SS is recognized again at the stage of addition of U4/U5/U6 snRNP to the spliceosome complex [33, 37]. The mechanism of this secondary recognition can be conveniently studied in the *trans*-splicing system since the two 5′SS recognition events can be easily separated. As mentioned above, spliceosome assembly in this system is stimulated by the addition of the 5′SS DNA competitor. Since excess of the 5′SS DNA (of the same sequence as the 5′SS RNA) competes with binding of the 5′SS RNA to U1 snRNP but not with spliceosome assembly, the latter process must specifically recognize some feature(s) of the 5′SS that are unique to the RNA substrate. Indeed, substitution of the uridine residue at position +2 of the intron (U+2, see Fig. 5) within the 5′SS RNA with thymidine

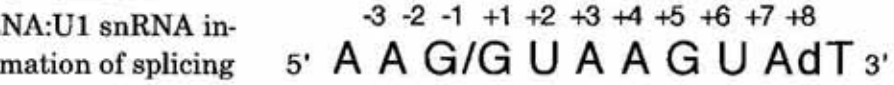

Figure 5. Sequence of the 5′ splice site RNA oligonucleotide used in the *trans*-splicing system.

(dT+2), results in a significant decrease in spliceosome assembly and splicing [38]. Interestingly, the observed requirement for the U+2 residue reflects an apparent recognition of the base, rather than the ribose backbone. Individual deoxy-substitutions along the entire 5′SS RNA sequence have only a minimal effect on the efficiency of spliceosome assembly and splicing [39]. Remarkably, even the 5′SS RNA containing only three ribonucleotides (G/GU, pos. −1, +1, +2) in the context of the otherwise DNA sequence, is recognized as a substrate and undergoes both steps of splicing [39]. Thus, the difference between the 5′SS RNA substrate for splicing and the 5′SS DNA must be restricted to just a few nucleotides around the splice junction and involves a specific recognition of the U+2 base.

The *trans*-splicing reaction illustrates several advantages offered by simplified, partial reaction systems. First, the use of a short RNA oligonucleotide as a splicing substrate bypasses the early interaction of the 5′SS

with U1 snRNP, simplifying the spliceosome assembly pathway. In general, by providing the conserved consensus 5′SS in the absence of any flanking sequences, the contribution of other RNA binding proteins that may interact with exon and intron sequences in the context of the full-length pre-mRNA *cis*-splicing substrate is minimized. This practically eliminates the application of any regulatory mechanisms used e.g. in alternative splicing reactions, but instead promotes efficient interactions with snRNP particles that are required for spliceosome assembly.

INTERACTION BETWEEN hPrp8 PROTEIN OF U5 snRNP AND THE CANONICAL GU DINUCLEOTIDE AT THE 5′ SPLICE SITE

Because of the highly efficient spliceosome assembly, the *trans*-splicing system was chosen to study specific interactions between the 5′SS and the spliceosomal components. Irradiation of splicing complexes with short-range (254 nm) UV light generates several characteristic crosslinks involving the 5′SS RNA. One of them represents the 5′SS RNA:U1 snRNA and is formed in the absence of the 5′SS DNA competitor in the reaction, when the 5′SS RNA binds directly to U1 snRNP [33]. Under *trans*-splicing conditions, i.e. in the presence of the 5′SS DNA and the 3′SS RNA, two other crosslinks are detected within splicing complex B: the 5′SS RNA:U6 snRNA and 5′SS RNA:hPrp8 [38]. The 5′SS RNA:U6 snRNA interaction represents the only RNA:RNA crosslink detected under these conditions, and corresponds to the previously identified base-pairing interaction between intron positions +4 to +6 and the highly conserved ACAGAG element in the U6 snRNA sequence in yeast and humans (see Fig. 2) [2–5, 37]. Potential base-pairing between the 5′SS and U6 snRNA involves intron positions +5 to +8, and is known to contribute to the stability of complex B and the overall splicing efficiency [40]. The presence of the 5′SS:U6 snRNA crosslink confirms the fidelity of the *trans*-splicing system and its utility to biochemical analysis of the spliceosome.

The second crosslink present in the splicing complex B, the 5′SS RNA:hPrp8p, represents the only 5′SS:protein interaction detected under these conditions [38]. A series of RNase digestions mapped the site of crosslink to the GU dinucleotide at the 5′ end of the intron, suggesting a close contact between hPrp8 and this highly conserved element within the 5′SS. Consistent with this notion, modification of the uridine (rU +2) of the GU to ribo thymidine (rT +2), which replaces a hydrogen (1.2 Å) at position 5 of the base with a methyl group (2.0 Å), results in strong inhibition of crosslinking to hPrp8, reduction in spliceosome assembly, and decrease in splicing catalysis [38]. This inhibitory effect reflects most likely a steric hindrance caused by a bulky group, since 5-iodoU (5IdU, 2.15 Å) modification at intron position +2 also shows similar negative effects. In contrast, introduction of a smaller 5-fluoroU (5FdU, 1.35 Å) group at the GU results in a 5′SS substrate similar to the unmodified RNA oligonucleotide. As expected, none of these uridine modifications had any effect when placed at position +6 in the intron [38]. These results strongly suggest that hPrp8 may specifically recognize the 5′SS GU dinucleotide and position it within the active site for the first step of splicing. A small subset of both U2- and U12-type pre-mRNAs is characterized by AU-AC (rather than GU-AG) dinucleotides at the ends of the intron [41, 42]. Since the same U5 snRNP, and thus the same Prp8, is implicated in splicing of both major and minor class of introns, the specificity of the 5′SS recognition by Prp8 may focus on the uridine (U+2), while the identity of the guanosine (G+1) may be recognized by different splicing factors. Further studies are needed to confirm this prediction.

THE HUMAN Prp8 PROTEIN – A CONSERVED COMPONENT OF U5 snRNP

The human Prp8 protein, a large (270 kDa), highly conserved component of U5 snRNP [43, 44], has been previously implicated in multiple contacts within the spliceosome. Site-specific crosslinking experiments demonstrated direct interactions of Prp8p with the 5′SS, branch site, polypyrimidine tract, and the 3′SS region in pre-mRNA [38, 45–49]. In fact, Prp8p represents the only known spliceosomal component implicated in direct interactions with all pre-mRNA signals important for splicing. In addition, as a stable component of U5 snRNP, it interacts with U5 snRNA and some other protein factors associated with this particle [50, 51]. Finally, genetic analysis in yeast suggests additional interactions between yPrp8 and other factors involved in splicing, including Prp40, DED1, and Prp28. Interestingly, the human ortholog of Prp28, U5-100 kDa, is an integral component of U5 snRNP [52], supporting the notion that Prp8 and Prp28 may interact directly with each other.

Because of the remarkable conservation of Prp8p, its apparently central position within the spliceosome, and its direct interaction with the conserved GU dinucleotide at the 5′ end of the intron, we were interested in determining the precise location of the 5′SS:hPrp8 crosslink. The mapping strategy involved proteolytic digestion of the purified 5′SS:hPrp8 crosslink with a series of proteases and chemical reagents. Knowing the specificity of individual cleavage reagents and the amino-acid sequence of hPrp8 protein, it was possible to limit the position of the crosslink to a short segment in the C-terminal portion of the protein. Furthermore, immunoprecipitation experiments using antibodies raised against various portions of the protein confirmed that the 5′SS RNA:hPrp8 crosslink resides in the C-terminal segment of this 2335 amino-acid protein [53].

Together with U6 snRNA, Prp8 represents one of the most highly conserved spliceosomal factors. This highly significant conservation spans a wide range of organisms – from humans (*H. sapiens*), through nematodes (*C. elegans*), yeast (*S. cerevisiae* and *S. pombe*), to trypanosomes (*T. brucei*) [43, 44]. For example, amino-acid sequences of *S. cerevisiae* and *C. elegans* Prp8 are 61.3% identical and 77.3% similar [43]. The protein sequence homology extends essentially over the entire sequence of Prp8, except for the N-terminal proline-rich segment (about 80 aa) found in *S. cerevisiae*, which is absent in the human or nematode protein. The high degree of sequence conservation allowed us to identify a region within *S. cerevisiae* Prp8 that corresponds to the mapped site of the 5′SS:hPrp8 crosslink [53]. Interestingly, the crosslink site is located C-terminally to one of the two most highly conserved regions in Prp8 (pos. 1889–1947 in yeast), which exhibits 93% identity between human and yeast sequences, while the region spanning the crosslink site itself is significantly less conserved (45% identity, pos. 1948–1979 in yeast). Since no identifiable protein motifs can be found in the Prp8p sequence, it is difficult to interpret the functional significance of these segments. However, the proximity of the crosslink site, indicating a direct physical contact with the 5′SS, suggests that the adjacent highly conserved region may play an important role in the reaction. The function of the conserved Prp8 segment could be limited to recognition of the 5′SS and its proper positioning at the catalytic center, but it could also include a direct participation in formation of the active site and perhaps in catalysis of splicing. Needless to say, additional studies are required to shed light on this important issue.

CONCLUSIONS

Naturally, recognition of the substrate RNA sequences represents a critical step in pre-

mRNA splicing. Recognition of the branch site seems relatively simple – it involves base pairing interaction between the branch site and U2 snRNA, facilitated by a number of RNA:protein interactions that allow and control binding of U2 snRNP to pre-mRNA (formation of splicing complex A). In contrast, recognition of the 5′SS appears to be more complex and involves at least two independent recognition events. First, the 5′SS is recognized by U1 snRNP at the stage of the commitment complex. To a large extent, this recognition relies on base pairing interaction between the 5′ end of U1 snRNA and the 5′SS consensus sequence. It is at this step in the reaction that most regulatory decisions concerning selection of the proper splice sites are made. The subsequent recognition of the 5′SS within the spliceosome involves an independent mechanism. The primary function of this recognition step is to precisely determine the phosphodiester bond that participates in splicing. The two recognition events are independent, as demonstrated by the *trans*-splicing system. It is remarkable that the 5′SS sequence alone, in the absence of the flanking sequences that typically control interactions with multiple RNA binding proteins, contains information sufficient to direct assembly of U2 and U4/U5/U6 snRNPs into a spliceosome-like complex. A number of specific interactions between the 5′SS and spliceosomal components must occur within this complex to precisely position the RNA in the catalytic center. It is this proximity of the splice site RNA substrate to the catalytic site of the complex that draws so much attention to detailed biochemical analysis of the 5′SS:spliceosome interactions. We hope that determination of precise contacts between the 5′SS substrate and the spliceosome will shed light on the structure of the catalytic center of this large and complex enzyme. In this context, the identified interaction between the most highly conserved GU dinucleotide at the 5′ end of the intron and the Prp8 factor of U5 snRNP is particularly interesting. While further studies are necessary to unravel the details of the mechanism by which Prp8 recognizes the 5′SS, these results strongly implicate Prp8 as an important component of the spliceosomal catalytic center. Of course, it is impossible to ignore the importance of the snRNA framework of the catalytic center; however, protein contribution should also be seriously considered. Among the spliceosomal proteins analyzed thus far, Prp8 is particularly interesting in this respect.

REFERENCES

1. Krämer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**, 367–409.
2. Moore, M.J., Query, C.C. & Sharp, P.A. (1993) Splicing of precursors to mRNA by the spliceosome; in *The RNA World* (Gesteland, R.F & Atkins, J.F., eds.) pp. 303–357, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
3. Nilsen, T.W. (1998) RNA-RNA interactions in nuclear pre-mRNA splicing; in *RNA Structure and Function* (Simons, R.W. & Grunberg-Manago, M., eds.) pp. 279–307, Cold Spring Harbor Laboratory Press, New York.
4. Nilsen, T.W. (1994) RNA-RNA interactions in the spliceosome: Unraveling the ties that bind. *Cell* **78**, 1–4.
5. Staley, J.P. & Guthrie, C. (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**, 315–326.
6. Hertel, K.J., Lynch, K.W. & Maniatis, T. (1997) Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* **9**, 350–357.
7. Newman, A. (1997) RNA splicing: Out of the loop. *Curr. Biol.* **7**, 418–420.
8. Wang, J. & Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa. *Curr. Opin. Genet. Dev.* **7**, 205–211.

9. Will, C.L. & Lührmann, R. (1997) Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320–328.

10. Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* **6**, 215–220.

11. Guthrie, C. (1994) The spliceosome is a dynamic ribonucleoprotein machine. *Harvey Lect.* **90**, 59–80.

12. O'Keefe, R.T., Norman, C. & Newman, A.J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86**, 679–689.

13. O'Keefe, R.T. & Newman, A.J. (1998) Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. *EMBO J.* **17**, 565–574.

14. Hall, S.L. & Padgett, R.A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357–365.

15. Hall, S.L. & Padgett, R.A. (1996) Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns (see comments). *Science* **271**, 1716–1718.

16. Tarn, W.-Y. & Steitz, J.A. (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* **84**, 801–811.

17. Tarn, W.Y. & Steitz, J.A. (1996) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns (see comments). *Science* **273**, 1824–1832.

18. Incorvaia, R. & Padgett, R.A. (1998) Base pairing with U6atac is required for 5′ splice site activation of U12-dependent introns *in vivo*. *RNA* **4**, 709–718.

19. Cech, T.R. (1983) RNA splicing: Three themes with variations. *Cell* **34**, 713–716.

20. Sharp, P.A. (1991) "Five easy pieces". *Science* **254**, 663.

21. Sharp, P.A. (1994) Split genes and RNA splicing. *Cell* **77**, 805–815.

22. Guthrie, C. (1991) Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science* **253**, 157–163.

23. Crispino, J.D., Blencowe, B.J. & Sharp, P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* **265**, 1866–1869.

24. Tarn, W.-Y. & Steitz, J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes & Dev.* **8**, 2704–2717.

25. Robberson, B.L., Cote, G.J. & Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94.

26. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414.

27. Abovich, N., Liao, X.C. & Rosbash, M. (1994) The yeast MUD2 protein: An interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes & Dev.* **8**, 843–854.

28. Legrain, P., Seraphin, B. & Rosbash, M. (1988) Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* **8**, 3755–3760.

29. Staknis, D. & Reed, R. (1994) SR proteins promote the first specific recognition of pre-mRNA and are present together with U1 small ribonucleoprotein in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**, 7670–7682.

30. Crispino, J.D., Mermoud, J.E., Lamond, A.I. & Sharp, P.A. (1996) *Cis*-acting elements distinct from the 5′ splice site promote U1-independent pre-mRNA splicing. *RNA* **2**, 664–673.

31. Alvarez, C.J., Romfo, C.M., Vanhoy, R.W., Porter, G.L. & Wise, J.A. (1996) Mutational analysis of U1 function in *Schizosaccharomyces pombe*: pre-mRNAs differ in the extent and

nature of their requirements for this snRNA *in vivo*. *RNA* **2**, 404–418.

32. Konforti, B.B. & Konarska, M.M. (1995) A short 5′ splice site RNA oligo can participate in both steps of splicing in mammalian extracts. *RNA* **1**, 815–827.

33. Konforti, B.B., Koziolkiewicz, M.J. & Konarska, M.M. (1993) Disruption of base pairing between the 5′ splice site and the 5′ end of U1 snRNA is required for spliceosome assembly. *Cell* **75**, 863–873.

34. Hall, K.B. & Konarska, M.M. (1992) The 5′ splice site consensus RNA oligonucleotide induces assembly of U2/U4/U5/U6 small nuclear ribonucleoprotein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10969–10973.

35. Maroney, P.A., Hannon, G.J., Shambaugh, J.D. & Nilsen, T.W. (1991) Intramolecular base pairing between the nematode spliced leader and its 5′ splice site is not essential for *trans*-splicing *in vitro*. *EMBO J.* **10**, 3869–3875.

36. Bruzik, J.P. & Steitz, J.A. (1990) Spliced leader RNA sequences can substitute for the essential 5′ end of U1 RNA during splicing in a mammalian *in vitro* system. *Cell* **62**, 889–899.

37. Madhani, H.D. & Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.* **28**, 1–26.

38. Reyes, J.L., Kois, P., Konforti, B.B. & Konarska, M.M. (1996) The canonical GU dinucleotide at the 5′ splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA* **2**, 213–225.

39. Reyes, J.L. (1998) Biochemical and genetic analysis of the interaction of Prp8 with the 5′ splice site during pre-mRNA splicing; in *Laboratory of Molecular Biology and Biochemistry*. The Rockefeller University, New York.

40. Crispino, J.D. & Sharp, P.A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes & Dev.* **9**, 2314–2323.

41. Dietrich, R.C., Incorvaia, R. & Padgett, R.A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell.* **1**, 151–160.

42. Sharp, P.A. & Burge, C.B. (1997) Classification of introns: U2-type or U12-type. *Cell* **91**, 875–879.

43. Hodges, P.E., Jackson, S.P., Brown, J.D. & Beggs, J.D. (1995) Extraordinary sequence conservation of the PRP8 splicing factor. *Yeast* **11**, 337–342.

44. Lücke, S., Klöckner, T., Palfi, Z., Boshart, M. & Binderief, A. (1997) *Trans* mRNA splicing in trypanosomes: Cloning and analysis of a *PRP8*-homologous gene from *Trypanosoma brucei* provides evidence for a U5-analogous RNP. *EMBO J.* **16**, 4433–4440.

45. Wyatt, J.R., Sontheimer, E.J. & Steitz, J.A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes & Dev.* **6**, 2542–2553.

46. Teigelkamp, S., Newman, A.J. & Beggs, J.D. (1995) Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**, 2602–2612.

47. Chiara, M.D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L. & Reed, R. (1996) Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol. Cell. Biol.* **16**, 3317–3326.

48. MacMillan, A.M., Query, Ch.C., Allerson, Ch.R., Chen, S., Verdine, G.L. & Sharp, P.A. (1994) Dynamic association of proteins with the pre-mRNA branch region. *Genes & Dev.* **8**, 3008–3020.

49. Umen, J.G. & Guthrie, C. (1995) A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes & Dev.* **9**, 855–868.

50. Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S. & Lührmann, R. (1998) The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol. Cell. Biol.* **18**, 6756–6766.

51. Dix, I., Russel, C.S., O'Keefe, R.T., Newman, A.J. & Beggs, J.D. (1998) Protein–RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4**, 1239–1250.

52. Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L. & Lührmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA* **3**, 1313–1326.

53. Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J. & Konarska, M.M. (1998) The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site. *RNA*, in press.