MOLECULAR AND CELLULAR BIOLOGY, July 1996, p. 3317–3326 Vol. 16, No. 7
0270-7306/96/$04.00+0

Identification of Proteins That Interact with Exon Sequences, Splice Sites, and the Branchpoint Sequence during Each Stage of Spliceosome Assembly

MARIA DOLORES CHIARA, OR GOZANI, MARIA BENNETT, PATRICK CHAMPION-ARNAUD, LEON PALANDJIAN, AND ROBIN REED*

Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115

Received 16 January 1996/Returned for modification 21 February 1996/Accepted 29 March 1996

We have carried out a systematic analysis of the proteins that interact with specific intron and exon sequences during each stage of mammalian spliceosome assembly. This was achieved by site-specifically labeling individual nucleotides within the 5′ and 3′ splice sites, the branchpoint sequence (BPS), or the exons with ^{32}P and identifying UV-cross-linked proteins in the E, A, B, or C spliceosomal complex. Significantly, two members of the SR family of splicing factors, which are known to promote E-complex assembly, cross-link within exon sequences to a region ~25 nucleotides upstream from the 5′ splice site. At the 5′ splice site, cross-linking of the U5 small nuclear ribonucleoprotein particle protein, $U5^{200}$, was detected in both the B and C complexes. As observed in yeast cells, $U5^{200}$ also cross-links to intron/exon sequences at the 3′ splice site in the C complex and may play a role in aligning the 5′ and 3′ exons for ligation. With label at the branch site, we detected three distinct proteins, designated BPS^{72}, BPS^{70}, and BPS^{56}, which replace one another in the E, A, and C complexes. Another dynamic exchange was detected with pre-mRNA labeled at the AG dinucleotide of the 3′ splice site. In this case, a protein, AG^{100}, cross-links in the A complex and is replaced by another protein, AG^{75}, in the C complex. The observation that these proteins are specifically associated with critical pre-mRNA sequence elements in functional complexes at different stages of spliceosome assembly implicates roles for these factors in key recognition events during the splicing pathway.

During pre-mRNA splicing, a series of highly dynamic spliceosomal complexes, consisting of multiple protein and small nuclear RNA (snRNA) components, assemble on pre-mRNA in the order E → A → B → C (for reviews, see references 5, 12, 20, 22, 27, and 31). The sequence elements required for assembly of these complexes are located at the 5′ and 3′ splice sites and at the branch site, and exon sequences affect the recognition of both splice sites. In metazoans, all of the elements involved in splicing are weakly conserved. Additional specificity is derived from strict constraints on the locations of these elements relative to one another, which presumably allows for complex networks of RNA-protein, protein-protein, and RNA-RNA interactions between factors bound to each element. Studies of the sequences required for splicing indicate that all of the critical elements are recognized multiple times during the splicing pathway. This proofreading, combined with the vast number of specific interactions established, is most likely the key to achieving high fidelity in the splicing reaction. A detailed understanding of the splicing mechanism requires identifying all of the factors that recognize each of the elements during the different stages of spliceosome assembly and understanding how these factors interact with both the pre-mRNA and one another.

A great deal of progress in identifying key recognition factors has come from a combination of genetic and biochemical analyses (for reviews, see references 5, 12, 20, 27, and 31). The 5′ splice site (consensus, RG/GURRGU [R = purine]) is first recognized by U1 small nuclear ribonucleoprotein particle (snRNP) in the E complex, and a duplex between U1 snRNA and the 5′ splice site is formed; the 5′ splice site is subsequently bound by U4/U5/U6 snRNP in the B complex, and U5 and U6 snRNAs base pair with exon and intron sequences, respectively (20, 27). The U5 snRNP protein, $U5^{200}$ (designated PRP8 in *Saccharomyces cerevisiae*), can also be specifically cross-linked to exon sequences at the 5′ splice site and may function to stabilize the U5 snRNA–pre-mRNA interaction (38, 44). Near or at the time of the second catalytic step of splicing, U5 snRNA and $U5^{200}$ interact with exon sequences at both the 5′ and 3′ splice sites, and this simultaneous interaction is thought to align the exons for ligation (38, 44).

The recognition element at the 3′ splice site consists of a 10- to 20-nucleotide (nt) pyrimidine tract followed by YAG (Y = pyrimidine). In metazoans, the pyrimidine tract is first recognized in the E complex by the splicing factor U2AF (45; reviewed in reference 16). During the E-to-A complex transition, U2AF is phosphorylated and becomes less tightly bound to pre-mRNA; U2AF remains loosely bound to the pyrimidine tract at least through B-complex assembly (9). The AG dinucleotide at the 3′ splice site is recognized at least twice during the metazoan splicing pathway. In introns containing a short and/or weak pyrimidine tract, the AG is required for catalytic step I of the splicing reaction, and in introns containing a strong pyrimidine tract, mutations of the AG affect the efficiency of this step (30, 33). Recognition of the AG is essential for catalytic step II in both types of introns. Factors that interact specifically with the AG have not been identified.

The BPS (branchpoint sequence; consensus, YNRAY [N = any nucleotide]) is essential for A-complex assembly, during which a short duplex is formed between the BPS and U2 snRNA (reviewed in reference 20). Stabilization of this U2 snRNA-BPS interaction appears to be largely mediated by the binding of several U2 snRNP proteins to a 20-nt region upstream of the BPS, designated the anchoring site (14). These U2 snRNP proteins are subunits of two multicomponent splicing factors, SF3a and SF3b, that are required for A-complex

* Corresponding author. Phone: (617) 432-2844. Fax: (617) 432-1144. Electronic mail address: rreed@warren.med.harvard.edu.

assembly (6, 7; see reference 16 for a review). The branch site adenosine, which is bulged from the U2 snRNA-BPS duplex, serves as the nucleophile for catalytic step I of the splicing reaction (28). Mutations of the branch site adenosine alone decrease A-complex assembly, indicating that this nucleotide is recognized prior to catalytic step I (reference 24 and this study). Site-specific labeling and UV cross-linking studies have identified a 14-kDa protein that cross-links to the branch site in the A complex and is a candidate for recognition of this site (29). The branch site adenosine is also required for catalytic step II of the splicing reaction, and a 28-kDa protein that cross-links to the BPS around the time of C-complex assembly may play a role in this recognition event (13). Several other proteins cross-link within a 15-Å (1.5-nm) radius of the BPS, but their identities and specific sites of cross-linking are not known (20).

Exon sequences play a central role in the use of both 5′ and 3′ splice sites and most likely first function during E-complex assembly. In many pre-mRNAs, specific purine-rich elements are present in exons and are required for excision of the intron immediately upstream (for reviews, see references 5, 12, and 31). Members of the SR family of splicing factors bind in a sequence-dependent manner to these purine-rich exonic enhancers and are required for their function (37; see reference 12 for a review). There are also some examples of non-purine-rich exonic enhancers that interact with SR proteins (36, 39, 40). It is not yet known whether the binding of SR protein family members to exons is a general phenomenon and, if so, whether these interactions account for the general effects of exon sequences on splicing.

In this study, we have used a site-specific ^{32}P labeling/UV cross-linking strategy to identify the proteins that interact with specific intron or exon sequences in each of the spliceosomal complexes (E, A, B, and C). Significantly, this analysis revealed that SR protein family members cross-link to exon sequences, and we mapped a major site of SR protein cross-linking to a region ~25 nt upstream of the 5′ splice site. Consistent with previous work, we found that the U5 snRNP protein, U5^{200}, initially interacts with sequences at the 5′ splice site and subsequently interacts with sequences at both the 5′ and 3′ splice sites. Finally, using pre-mRNAs labeled at the BPS or at the AG dinucleotide of the 3′ splice site, we detected multiple dynamic exchanges of novel cross-linked proteins. The functional significance of all of these RNA-protein interactions is indicated by the observations that they are detected only at specific sites in the pre-mRNA, only at specific stages of spliceosome assembly, and only in functional spliceosomal complexes. Previous mutational analyses of the branch site and AG dinucleotide indicated that these sequence elements are recognized multiple times during the splicing pathway. The cross-linked proteins that we have identified are excellent candidates for factors involved in these recognition events.

MATERIALS AND METHODS

Plasmids. The plasmid encoding wild-type (WT) adenovirus major late (AdML) pre-mRNA is pAdML (21), the plasmid encoding GG pre-mRNA is pAdMLΔAG (15), and the plasmid encoding ΔBS pre-mRNA is pΔBS (8). WT and ΔBS pre-mRNAs are identical except that ΔBS contains an A-to-C substitution of the branch site adenosine. WT and GG pre-mRNAs are identical except that GG contains an A-to-G substitution in the AG dinucleotide at the 3′ splice site and a strengthened pyrimidine tract. The 3′ splice site sequence for WT pre-mRNA is cctgtccctttttttttccacag, and that for GG is cctttccctttttttttcc tctctctctgg. DNAs were linearized with *Bam*HI for in vitro transcription and transcribed with T7 RNA polymerase.

Synthesis of site-specifically labeled pre-mRNA. Site-specifically ^{32}P-labeled AdML pre-mRNAs were synthesized as described previously (23). DNA templates encoding the appropriate 5′ and 3′ portions of the transcript were generated by PCR, fractionated on an agarose gel, and purified by using a Qiaex II gel extraction kit (Qiagen). The 5′ portion of the transcript was then synthesized in a standard T7 RNA polymerase transcription reaction mixture containing final concentrations of 400 μM each ATP and CTP, 200 μM each GTP and UTP, and 840 μM 5′ cap nucleotide, GpppG. For synthesis of the 3′ portion of the transcript, guanosine (at a concentration 10-fold greater than that of GTP) was included in the transcription reaction mixture, and the 5′ cap nucleotide was omitted. 5′ and 3′ transcripts were purified by fractionation on 6.5% denaturing polyacrylamide gels, followed by elution from the gel and ethanol precipitation. The 3′ portion of the transcript (50 pmol) was then 5′ end labeled in a 20-μl reaction mixture containing 2.2 μl of [γ-^{32}P]ATP (7,000 Ci/mmol), 2 μl of T4 polynucleotide kinase, and 4 μl of OPA buffer (Pharmacia). The 5′ and 3′ portions of the transcript were ligated in 50-μl reaction mixtures containing 1.7 μl of T4 DNA ligase (2×10^6 U/ml; New England Biolabs), 1 μl of RNA Guard (Pharmacia), 5 μl of DNA ligation buffer (New England Biolabs), 75 pmol of the 5′ portion of the transcript, 50 pmol of the 3′ portion of the transcript, and 50 pmol of a 25- to 30-nt DNA-bridging oligonucleotide (23). Ligations were carried out at 25°C for 3 to 12 h. Guanosine residues were used for site-specific labeling because transcription initiates most efficiently with this nucleotide. For region-specifically ^{32}P-labeled pre-mRNAs, the labeled portion of the RNA was synthesized in 100-μl reaction mixtures containing 6.25 μl each of ^{32}P-labeled GTP, UTP, CTP, and ATP (3,000 Ci/mmol). Unlabeled nucleoside triphosphates were present in the reaction mixture at a concentration of 40 μM each. Reaction mixtures also contained 400 μM cap (for synthesis of the 5′ portion of the transcript) or 400 μM GMP (for synthesis of the 3′ portion of the transcript).

Isolation of spliceosomal complexes and UV cross-linking. Spliceosomal complexes A, B, and C were assembled by incubating 600 ng of pre-mRNA (600 ng of unlabeled pre-mRNA plus the site-specifically labeled pre-mRNA, which was present in negligible amounts) in a 0.75-ml standard splicing reaction mixture for 5, 10, and 30 min, respectively. Assembly of the E complex was performed as described previously (21). Complexes were isolated by gel filtration, and 200-μl aliquots were cross-linked by irradiation with 250-nm UV light for 5 min at 4°C at a distance of 5.5 cm from the light source (Sylvania G15T8 lamp) (9, 36). RNase A (4 to 10 μg), or where indicated RNase T_1 (4 to 10 μg), was added to 150-μl aliquots of the gel filtration fractions and incubated for 30 min at 37 or 45°C, respectively. Proteins were analyzed on sodium dodecyl sulfate (SDS)-containing or two-dimensional (2D) gels. Native gel electrophoresis of splicing complexes was carried out as described previously (18).

RESULTS

To identify proteins that interact with critical sequence elements during spliceosome assembly, we synthesized AdML pre-mRNAs containing a single ^{32}P-labeled nucleotide. These site-specifically labeled pre-mRNAs were assembled into spliceosomal complexes, isolated by gel filtration, and UV irradiated (see Materials and Methods) (14). Cross-linked complexes were then treated with RNase A or RNase T_1 (these enzymes cleave RNA after pyrimidine or guanosine residues, respectively), and the proteins were fractionated on SDS-containing or 2D gels. In this technique, cross-linked proteins are not mapped to an individual nucleotide in the pre-mRNA but are instead localized to one or more RNase fragments in the immediate vicinity of the labeled nucleotide. In some cases, the bound proteins partially protect the RNA from RNase digestion, making it difficult to conclusively localize the cross-linked protein to a single RNase digestion product. Thus, throughout our study, we indicate the nucleotides that are labeled but only are able to localize the cross-linked proteins to the vicinity of these sites. The A, B, and C complexes used in our study were obtained from splicing reactions incubated for 5, 10, and 30 min, respectively. On WT pre-mRNA, maximal levels of C complex (which contains exon 1 and the lariat intermediate) are present at the 30-min time point (data not shown).

Initially, we analyzed the cross-linked proteins detected with pre-mRNA site-specifically labeled at a position 3 nt upstream from the 5′ splice junction (E^{-3} [Fig. 1A]). Cross-linked proteins detected in the H, A, B, and C complexes were compared (Fig. 1B). Throughout our study, equal counts per minute were loaded in the lanes shown in a given panel, and thus the levels of the cross-linked proteins in each panel can be compared directly. At the E^{-3} site, we detected a ~200-kDa protein that cross-links in the B complex (designated U5^{200}; see below) (Fig. 1B, lane 3). This protein does not cross-link in the het-

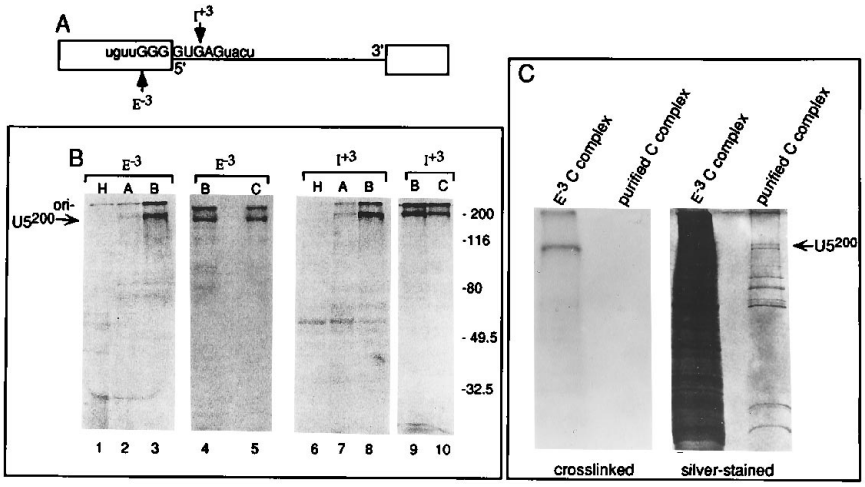


FIG. 1. $U5^{200}$ cross-links to the 5′ splice site in the B and C complexes. (A) Schematic of AdML pre-mRNA showing the sequence in the vicinity of the 5′ splice site (the 5′ splice site sequence element is in capitals) and the positions in the exon and intron that were site specifically labeled. (B) The indicated spliceosomal complexes were assembled on site-specifically ^{32}P-labeled pre-mRNA, isolated by gel filtration, UV cross-linked, and then treated with RNase A. In each panel, equal counts per minute were loaded on an SDS–9% polyacrylamide gel. $U5^{200}$, the molecular mass markers (in kilodaltons), and the origin (ori) of the gels are indicated. (C) Cross-linked C complex assembled on pre-mRNA labeled at the E^{-3} site was fractionated next to purified C complex on an SDS–7.5% polyacrylamide gel. The silver-stained pattern (right panel) and autoradiogram (left panel) are shown. The silver-stained pattern of the cross-linked complex (E^{-3} C complex) is very dark because total proteins from the gel filtration fraction are present in the sample.

erogeneous nuclear ribonucleoprotein particle (hnRNP) complex H (lane 1) and cross-links at very low levels in the A complex (lane 2). A comparison of the B and C complexes revealed similar levels of cross-linking of the 200-kDa protein in these two complexes (Fig. 1B, lanes 4 and 5). We obtained the same results when pre-mRNA was labeled 3 nt downstream of the 5′ splice junction (I^{+3} [Fig. 1B, lanes 6 to 10]). On the basis of these data, we conclude that a 200-kDa protein cross-links in the vicinity of the 5′ splice site. As noted above, our data do not reveal the precise cross-linking site of this protein. If the pre-mRNA in this region was digested to completion by the RNase A treatment, our data would indicate that there are two cross-linking sites for the 200-kDa protein. The two cross-linked adducts would be present on the U residues immediately 5′ to the labeled E^{-3} and I^{+3} nucleotides (Fig. 1A; note that RNase A digestion generates 3′ phosphates). Alternatively, if the pre-mRNA in this region is partially protected from RNase A digestion by the bound protein, there may be only one cross-linking site that is detected by labeling at either the E^{-3} or the I^{+3} site.

Previous work showed that the U5 snRNP protein, $U5^{200}$, is present in spliceosomal complexes B (4) and C (15) and that this protein cross-links to pre-mRNA in both of these complexes (15, 35). In addition, site-specific labeling studies showed that $U5^{200}$ cross-links at the 5′ splice site in the spliceosome (38, 44). To determine whether the 200-kDa protein that we detect cross-linking at the 5′ splice site corresponds to $U5^{200}$, we fractionated affinity-purified spliceosomal complex C, which contains $U5^{200}$ (15), side by side with our cross-linked sample (Fig. 1C). Superimposing the autoradiogram of the cross-linked sample and the silver-stained C complex revealed that the cross-linked protein comigrates with $U5^{200}$ (the top band of the doublet; we note that the cross-linked sample silver stains very darkly because it contains total protein from the gel filtration fraction). These data, together with the previous work mentioned above, indicate that the 200-kDa protein that we detect at the 5′ splice site is $U5^{200}$.

In yeast cells, $U5^{200}$ cross-links to exon and intron sequences at the 3′ splice site around the time of catalytic step I of the splicing reaction (38, 41). To determine whether these interactions also occur in mammals, we analyzed the cross-linked proteins on pre-mRNA site-specifically labeled at the E^{+7} position at the 3′ splice site (Fig. 2A). Consistent with the results for *S. cerevisiae*, a 200-kDa cross-linked protein is detected at this site in the C complex but not in the H, E, A, or B complex (Fig. 2B and data not shown). This 200-kDa protein likely corresponds to $U5^{200}$, as these proteins comigrate on SDS-containing and 2D gels (data not shown).

Cross-linking of $U5^{200}$ can also be detected with label at the AG dinucleotide (I^{-1} in the C complex [Fig. 2C, lane 3; the cross-linked band below the 87-kDa marker is described below]). In yeast cells, mutation of the AG dinucleotide does not prevent cross-linking of $U5^{200}$ (38, 41). To determine whether mutation of the AG affects cross-linking of $U5^{200}$ in mammals, we analyzed GG pre-mRNA (15). This pre-mRNA contains an A-to-G substitution in the AG dinucleotide and a strengthened pyrimidine tract, which allows an efficient catalytic step I in the absence of step II (see Materials and Methods for sequence comparison of WT and GG pyrimidine tracts). As shown in Fig. 2C, $U5^{200}$ cross-linking is detected at the I^{-1} site with GG pre-mRNA. Thus, in mammals as in *S. cerevisiae*, $U5^{200}$ interacts with the 3′ splice site after catalytic step I of the splicing reaction (38, 41).

SR protein family members cross-link to exon 1 sequences. Members of the SR family of splicing factors promote the binding of U1 snRNP to the 5′ splice site and U2AF to the 3′ splice site during assembly of the E complex (11, 17, 36; for reviews, see references 12 and 31). Purified SR proteins bind to 5′ splice site sequences (46) as well as to purine-rich motifs found in exonic enhancers (37; reviewed in reference 12). Previously, we showed that two SR protein family members, SRp 20 and SRp 30, UV cross-link to AdML pre-mRNA in the E and B spliceosomal complexes (36). In the present study, we did not detect these SR proteins cross-linking at either the 5′ or the 3′ splice site (Fig. 1 and 2). Thus, to determine where the SR proteins do cross-link, we analyzed the cross-linked proteins in the E complex by using region-specifically labeled pre-mRNAs. These pre-mRNAs contained fully labeled exon

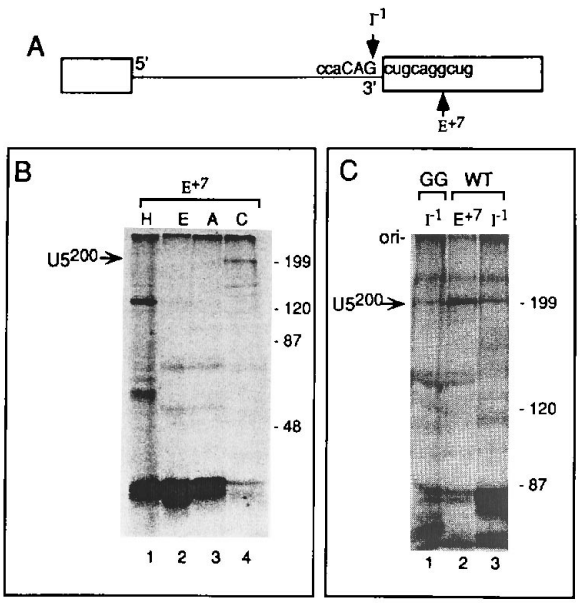


FIG. 2. $U5^{200}$ cross-links to intron/exon sequences at the 3′ splice site in the C complex. (A) Schematic of AdML pre-mRNA showing the sequence of the 3′ splice site and 5′ end of exon 2 (the conserved CAG at the 3′ splice site is in capitals). The positions that were site-specifically ^{32}P labeled are shown. (B) Spliceosomal complexes were assembled on site-specifically labeled pre-mRNAs, UV cross-linked, treated with RNase A, and fractionated on SDS–9% polyacrylamide gels. Equal counts per minute were loaded in each lane. Molecular mass markers (in kilodaltons) and $U5^{200}$ are indicated. (C) Same as panel B except that the C complex was assembled on GG or WT pre-mRNA, and proteins were fractionated on a 7.5% gel. The band above $U5^{200}$ in all three lanes was not detected reproducibly.

1, exon 1 plus intron 1, or exon 2 (schematic in Fig. 3A; labeled regions are indicated in black). As shown previously (36), SRp 20, SRp 30, and $U2AF^{65}$ cross-link strongly in the E complex but not in the H complex, whereas the reverse is true for hnRNP I (Fig. 3A; compare lanes 1 to 3 with lanes 4 to 6). All of these proteins migrate on a 2D gel in their expected positions (36) (Fig. 3B). Analysis of the region-specifically labeled pre-mRNAs shows that SRp 20 and SRp 30 cross-linking is strongly detected when exon 1 plus intron 1 or exon 1 alone is labeled, whereas SRp 30 is weakly detected when exon 2 is labeled (Fig. 3A, lanes 4 to 6). These observations indicate that the SR proteins largely cross-link within exon 1. We can deduce that little SR protein cross-linking occurs on the intron because the total amount of SR protein cross-linking detected on exon 1 alone is roughly the same as that detected on the whole pre-mRNA (exon 1 plus intron 1 and exon 2, combined [Fig. 3A, lanes 4 and 6]).

To more precisely map the cross-linking sites of the SR proteins within exon 1, we used site-specifically labeled pre-mRNAs (Fig. 4A). As shown above (Fig. 1B), no SR protein cross-linking is detected at the E^{-3} position in the A, B, or C complex. Similarly, we do not detect SR proteins cross-linking at this site in the E complex (data not shown). In contrast, SRp 20 and SRp 30 cross-linking is detected at the E^{-26} site but not at the E^{-15} site (Fig. 4B, E^{-15} and E^{-26}). On the basis of this observation, we labeled pre-mRNAs at the E^{-19}, E^{-26}, and E^{-31} sites (Fig. 4A) and compared cross-linked complexes treated with RNase A or RNase T_1 (Fig. 4C). With RNase A, SRp 20 and SRp 30 are detected only at the E^{-26} site. In contrast, with RNase T_1, SRp 20 and SRp 30 are detected at both the E^{-26} and E^{-31} sites (Fig. 4C). In addition, another protein, which corresponds to $U2AF^{65}$ (Fig. 3B and legend), is detected on the E^{-31} RNase T_1 fragment. This RNase T_1 fragment is large (16 nt; underlined in Fig. 4A), which explains the increase in apparent molecular weight of the cross-linked proteins (Fig. 4C; compare sizes of SR proteins in RNase T_1, E^{-26}, and E^{-31} lanes). In addition, the E^{-31} RNase T_1 fragment is pyrimidine rich, which presumably accounts for the cross-linking of U2AF at this site. Together, the data indicate that there is more than one cross-linking site for SRp 20 and SRp 30 in this region. One cross-linking site for these proteins

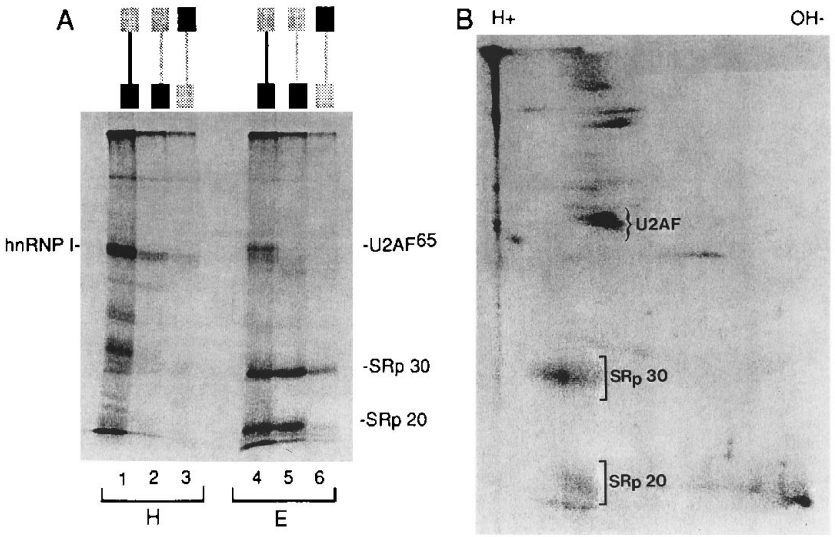


FIG. 3. SR proteins cross-link to exon sequences upstream of the 5′ splice site. (A) AdML pre-mRNAs were region-specifically ^{32}P labeled as follows. Exon 1 was labeled from the 5′ terminus of the pre-mRNA to the E^{-3} position, exon 1 plus intron 1 was labeled from the 5′ terminus to the I^{-1} site, and exon 2 was labeled from the I^{-1} site to the 3′ terminus of the pre-mRNA (labeled portions are indicated in black in the schematics) (see Fig. 1A and 2A for sequences at labeling sites). Site-specifically labeled pre-mRNAs were assembled into a spliceosomal complex (H or E), UV cross-linked, treated with RNase A, and fractionated on an SDS–9% polyacrylamide gel. The SR proteins, U2AF, and hnRNP I are indicated. (B) Cross-linked E complex assembled on region-specifically labeled pre-mRNA (labeled from the 5′ end of exon 1 to the E^{-26} site; see Fig. 4B) was fractionated on a 2D gel. The SR proteins and U2AF are indicated. The bands corresponding to SR proteins and U2AF were identified previously (36).

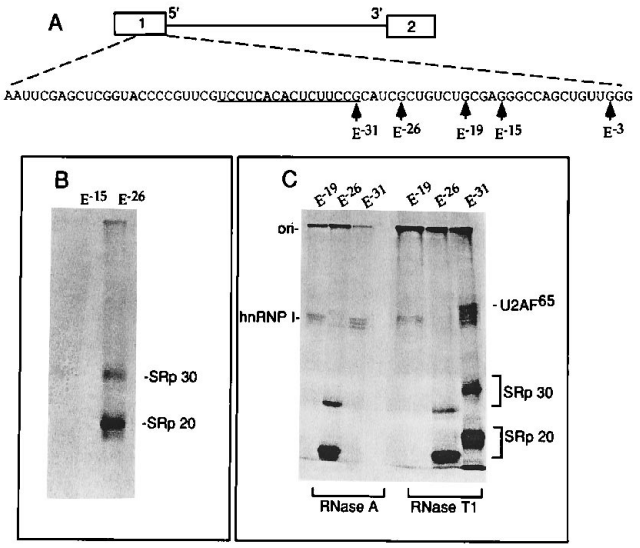


FIG. 4. SR proteins cross-link to a region 25 nt upstream from the 5′ splice site. (A) Schematic of AdML pre-mRNA. The sequence of a portion of exon 1 is shown, and the sites that were ^{32}P labeled are indicated. The RNase T_1 fragment detected when pre-mRNA is labeled at the −31 site is underlined. (B and C) Spliceosomal complex E was assembled on AdML pre-mRNA labeled at the sites indicated. Complexes were UV cross-linked, treated with RNase A or T_1 as indicated, and fractionated on SDS–9% polyacrylamide gels. The origin of the gel (ori), SR proteins, U2AF and hnRNP I are indicated.

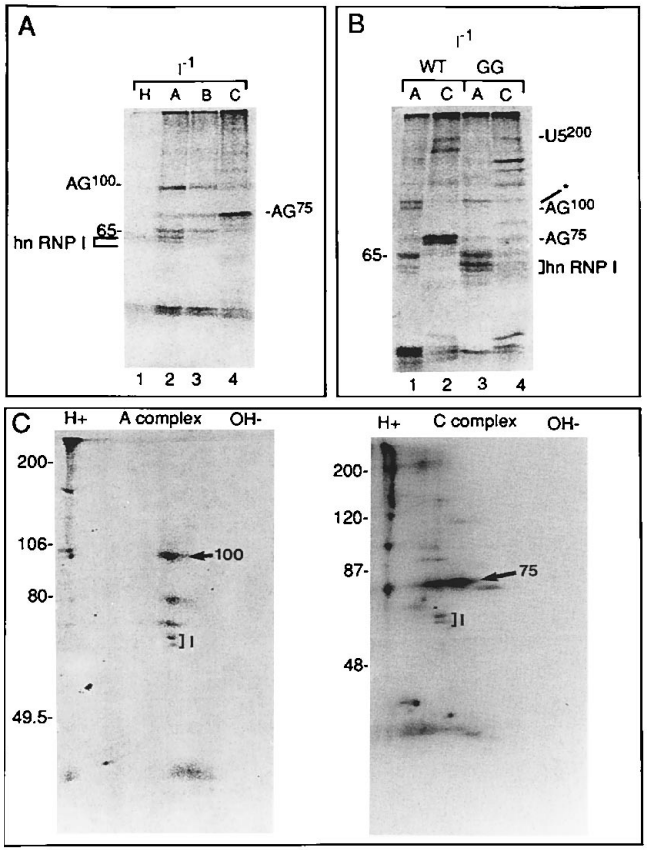


FIG. 5. Proteins of 100 and 75 kDa cross-link to the AG dinucleotide in the A/B and C complexes, respectively. (A) AdML pre-mRNA site-specifically ^{32}P labeled at the guanosine residue in the AG dinucleotide at the 3′ splice site was assembled into the indicated spliceosomal complexes, UV cross-linked, treated with RNase A, and fractionated on an SDS–9% polyacrylamide gel. Note that less H complex was loaded on this gel. In other experiments, hnRNP I and the low-molecular-weight doublet seen in lanes 2 to 4 is detected in the H complex. (B) Cross-linked A or C complex assembled on WT pre-mRNA or GG pre-mRNA (containing an A-to-G mutation in the AG dinucleotide) and labeled at the G residue in the AG dinucleotide was fractionated on an SDS–9% polyacrylamide gel. Note that the band above AG^{100} indicated with the asterisk is not reproducibly detected and was present in the H complex in this experiment (data not shown). (C) Same as panel A except that complexes (A and C) were fractionated on 2D gels. Molecular mass markers (in kilodaltons), hnRNP I, and AG-specific cross-linked proteins are indicated.

is on the E^{-26} RNase T_1 fragment, while another site for these proteins, together with $U2AF^{65}$, is on the E^{-31} RNase T_1 fragment. However, from these data we cannot distinguish between the possibilities that more than one SR protein is present on every pre-mRNA molecule and that there are different populations of pre-mRNA containing different combinations of these proteins. In any case, our data show that a specific region in exon 1 located about 25 nt upstream from the 5′ splice site is a major site of SR protein cross-linking in AdML pre-mRNA. Further studies are needed to determine whether there are additional cross-linking sites for these proteins upstream that were not detected in our site-specific labeling analysis. We note that although $U2AF^{65}$ cross-linking to exon 1 is readily detected by site-specific labeling (e.g., Fig. 4C), its level of cross-linking in exon 1 (detected with region-specifically labeled pre-mRNA) is much less than in exon 1 plus intron 1 (Fig. 3A, lanes 4 and 5 and longer exposures of these lanes [data not shown]). Thus, these data are consistent with previous work indicating that the major site of $U2AF^{65}$ cross-linking is at the 3′ splice site (45).

Sequential interactions of two distinct proteins at the 3′ splice site. To identify candidate factors for recognition of the essential AG dinucleotide at the 3′ splice site, we assembled spliceosomal complexes on pre-mRNA labeled at the guanosine residue in the AG (I^{-1} site) and carried out UV cross-linking. No specific cross-linked protein was detected with label at this site in the E complex (data not shown). In contrast, cross-linking of a 100-kDa protein, which we have designated AG^{100}, was detected in the A complex (Fig. 5A, lane 2). This protein is not detected in the H complex (Fig. 5A, lane 1), even on long exposures (data not shown; see the legend to Fig. 5A). The levels of AG^{100} cross-linking decrease as spliceosome assembly proceeds, and a new protein of 75 kDa, AG^{75}, replaces AG^{100} in the C complex (Fig. 5A; compare lanes 2 to 4).

Assuming that the pre-mRNA in the vicinity of the AG has been digested to completion, our data indicate that AG^{75} and AG^{100} are present on the 3-nt RNase A fragment (AGC) that contains the AG dinucleotide (see Fig. 2A for sequence of the 3′ splice site). If incomplete RNase digestion has occurred, then these proteins may cross-link further upstream or downstream of the AG. However, as shown above, neither AG^{75} nor AG^{100} is detected at the E^{+7} site in the exon (Fig. 2). In addition, these proteins are not detected in the intron at a position 6 nt upstream of the AG (data not shown). Thus, we conclude that AG^{75} and AG^{100} interact with pre-mRNA very close to, if not at, the AG dinucleotide.

To determine whether the AG dinucleotide plays a role in the binding of AG^{75} or AG^{100} to pre-mRNA, we analyzed GG pre-mRNA site-specifically labeled at the I^{-1} site (15) (see Materials and Methods for sequences of WT and GG 3′ splice sites). As shown in Fig. 5B, cross-linking of both AG^{75} and AG^{100} was dependent on the AG. In contrast, a ~65-kDa protein, which first cross-links to WT pre-mRNA in the A complex and then diminishes (Fig. 5A; compare lanes 2 to 4),

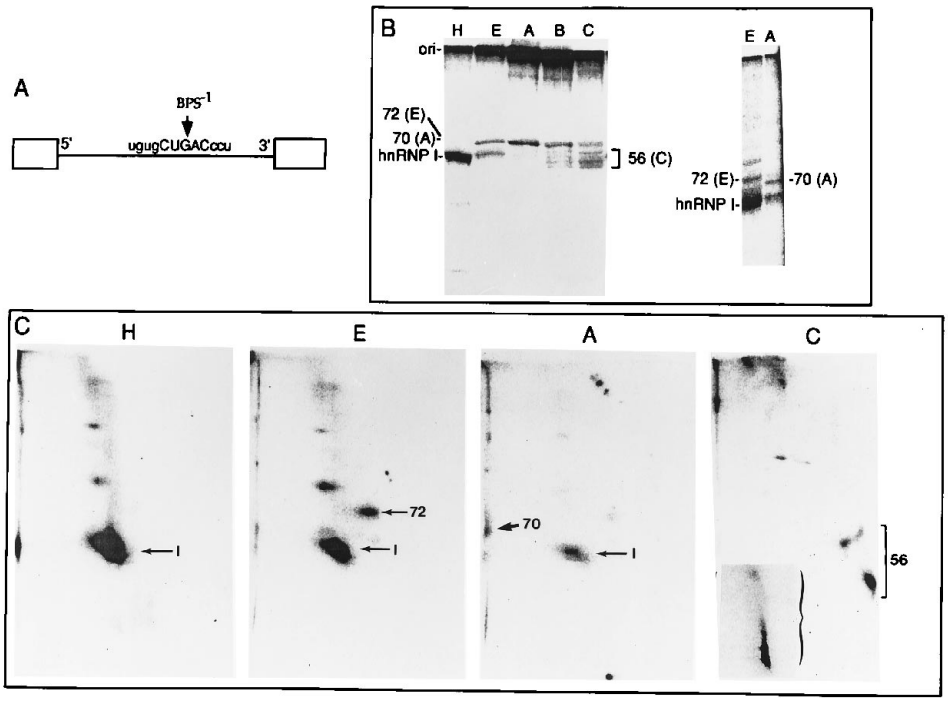


FIG. 6. Proteins of 70, 72, and 56 kDa cross-link to the branch site in the E, A/B, and C complexes, respectively. (A) Schematic of AdML pre-mRNA showing the sequence in the vicinity of the BPS (the BPS element is shown in capitals). The site that was ^{32}P labeled is indicated. (B) Pre-mRNA labeled at the BPS^{-1} site was assembled into the indicated complexes, UV cross-linked, treated with RNase A, and fractionated on an SDS–9% polyacrylamide gel. The origin of the gel is indicated (ori). In the right panel, the gel was run longer to separate BPS^{70} and BPS^{72}. (C) Same as panel B except that the indicated complexes were fractionated on 2D gels. The acidic end of the gels is on the left. hnRNP I (I) and the BPS-specific proteins are indicated. An inset showing a longer exposure of BPS^{56} is shown in panel C. Sizes are indicated in kilodaltons.

still cross-links to GG pre-mRNA (Fig. 5B, lanes 1 and 3). Similarly, as observed above (Fig. 2C), $U5^{200}$ cross-links normally to GG pre-mRNA (Fig. 5B; compare lanes 2 and 4; note that $U5^{200}$ cross-linking at this site is sometimes difficult to detect [e.g., Fig. 5A, lane 4]). Thus, these data are consistent with the possibility that the AG is required for the binding of AG^{75} and AG^{100} but not for the binding of $U5^{200}$ and the 65-kDa protein. However, in the case of AG^{75} and AG^{100}, we cannot rule out the possibility that the binding of these proteins to GG pre-mRNA occurs normally, but the cross-linking is disrupted due to the sequence change.

To determine whether AG^{100} and AG^{75} correspond to proteins previously identified in purified spliceosomal complexes, we fractionated the I^{-1} cross-linked A and C complexes on 2D gels (Fig. 5C). Comparison of these cross-linking patterns with those of purified A or C complexes indicates that AG^{100} and AG^{75} do not correspond to any proteins previously detected by cross-linking or by silver staining (data not shown) (15, 35). We conclude that AG^{100} and AG^{75} are novel spliceosomal proteins.

Three distinct proteins interact sequentially with the BPS. To identify potential branch site recognition factors, pre-mRNA was labeled 1 nt upstream of the branch site adenosine (BPS^{-1} [Fig. 6A]). Several cross-linked proteins that range from 50 to 70 kDa are detected at this site during spliceosome assembly (Fig. 6B, left panel). To characterize these proteins further, the cross-linked spliceosomal complexes were fractionated on 2D gels (Fig. 6C). In the H complex, hnRNP I is the main cross-linked protein (Fig. 6B, left panel; Fig. 6C, panel H). In this experiment, the H complex contaminated the E and A complexes, which accounts for the strong hnRNP I cross-linking in these complexes (Fig. 6C, panels E and A, and data not shown). In the E complex, a protein designated BPS^{72} is detected (Fig. 6B, left panel; Fig. 6C, panel E). Analysis on 2D gels reveals that BPS^{72} cross-linking in the E complex is replaced by cross-linking of another protein, BPS^{70}, in the A complex (Fig. 6C; compare panels E and A). In contrast to most spliceosomal proteins, BPS^{70} does not enter the isofocusing dimension of our 2D gels, most likely because it is very acidic (Fig. 6C, panel A). An SDS gel in which BPS^{72} and BPS^{70} are resolved is shown in Fig. 6B (right panel). By the time that the C complex assembles, BPS^{72} cross-linking has diminished, and a protein designated BPS^{56} cross-links (Fig. 6B, left panel; compare lanes A, B, and C). This protein fractionates as a heterogeneous band, indicating that it may be partially proteolyzed and/or modified (e.g., Fig. 6C, panel C; the inset shows a darker exposure of BPS^{56}). None of the proteins detected with pre-mRNA labeled at the branch site appear to correspond to cross-linked proteins previously identified on 2D gels of spliceosomal complexes (data not shown) (15, 35). We conclude that three novel proteins, BPS^{72}, BPS^{70}, and BPS^{56}, replace one another at or very near the branch site in the E, A, and C complexes. Assuming that the RNase A digestion of the cross-linked complexes was complete, the U residue immediately upstream of the labeled G would contain the cross-linked adduct (Fig. 6A). If the RNase digestion was not complete, we know that the cross-linking site must be within 6 nt upstream and 5 nt downstream of the branch site adenosine. This conclusion is based on the observation that other cross-linked proteins are detected at these sites, and the BPS proteins are not detected (14) (data not shown). In previous work, proteins of 14 kDa (29) and 28 kDa (13) were detected in the A and C complexes, respectively. On high-percentage gels, we were unable to detect cross-linked species

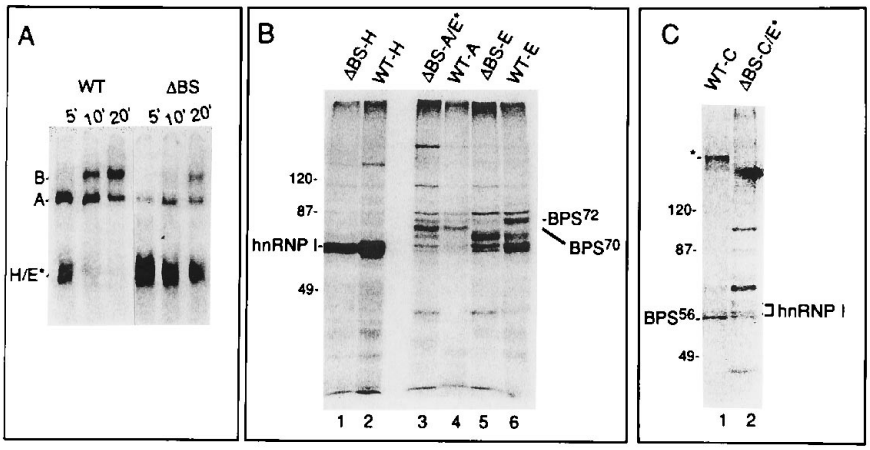


FIG. 7. Effect of branch site mutation on proteins cross-linking at the BPS. (A) WT and ΔBS pre-mRNAs were incubated under splicing conditions for the times indicated (in minutes) and fractionated on a native gel. The bands corresponding to the E*, H, A, and B complexes are shown. (B and C) WT and ΔBS pre-mRNAs site-specifically labeled at the BPS^{-1} site were assembled into the indicated spliceosomal complexes, UV cross-linked, treated with RNase A, and fractionated on 9% (B) or 7.5% (C) polyacrylamide gels. Molecular mass markers (in kilodaltons), hnRNP I, and BPS-specific proteins are indicated.

in these size ranges (data not shown). These discrepancies are likely due to the differences in cross-linking methods and/or nucleotides labeled (see Discussion).

Mutation of the BPS does not have any obvious effects on the efficiency of E-complex assembly (8). In contrast, a pre-mRNA which contains an A-to-C substitution of the branch site adenosine (ΔBS pre-mRNA) does not assemble A and B complexes efficiently (24) (Fig. 7A). Previous work showed that a novel ATP-dependent complex, designated the E* complex, accumulates on ΔBS and other branch site mutants; this complex is detected in the position of the spliceosomal complexes by gel filtration but cofractionates with the H complex on native gels (8). The E* complex lacks stably bound U2 snRNP but, unlike the E complex, does not contain tightly bound U1 snRNP and U2AF. To determine whether BPS^{72}, BPS^{70}, or BPS^{56} cross-linking is affected by branch site mutations, we site-specifically labeled ΔBS pre-mRNA at the BPS^{-1} site and analyzed the cross-linked proteins in the E complex, the A/E* complex, or the C/E* complex, respectively (note that both the E* and normal spliceosomal complexes assemble on ΔBS pre-mRNA, apparently as a result of some use of the mutant branch site [8] [Fig. 7A]).

As shown in Fig. 7B, BPS^{72} cross-linking in the E complex is abolished in ΔBS pre-mRNA (compare lanes 5 and 6). In contrast, BPS^{70} cross-linking in the A/E* complex appears to be increased in the BPS mutant, but further work is needed to determine whether these proteins are one and the same (Fig. 7B; compare lanes 3 and 4). Finally, BPS^{56} cross-linking is abolished in the C/E* complex (Fig. 7C; compare lanes 1 and 2); this result was confirmed on 2D gels which separate BPS^{56} from the hnRNP I doublet (data not shown). We conclude that the branch site adenosine is required for cross-linking of BPS^{72} and BPS^{56} at the time of E- and C-complex assembly, respectively.

DISCUSSION

Spliceosome assembly involves the formation of highly complex and specific networks of interactions between proteins, snRNAs, and the pre-mRNA, and these interactions are established and disrupted during the different stages of spliceosome assembly. The recognition elements that nucleate assembly of the networks are located at the 5′ and 3′ splice sites, at the branch site, and in the exons. To identify key factors involved in recognizing these elements, we used a site-specific ^{32}P labeling and UV cross-linking strategy to detect proteins in each of the spliceosomal complexes (E, A, B, and C). Our data are summarized in the schematics shown in Fig. 8. In the E complex, SR protein family members cross-link to AdML pre-mRNA within exon 1, and we identified a region about 25 nt upstream of the 5′ splice site as a major cross-linking site. The SR proteins also cross-link to pre-mRNA in the B complex (36) and thus may remain bound to the exon throughout spliceosome assembly. The other specific interaction that we detect in the E complex is cross-linking of a protein, BPS^{72}, to a site near or at the BPS. No specific cross-linked proteins were detected in the E complex with pre-mRNA labeled at either the 5′ splice site or the AG dinucleotide of the 3′ splice site. In the A complex, BPS^{72} is replaced by BPS^{70}, and AG^{100} first cross-links to a site at or near the AG dinucleotide. In the B complex, cross-linking at the 5′ splice site is first detected, and our data are consistent with previous work identifying the cross-linked protein at this site as the U5 snRNP protein, $U5^{200}$ (38, 44). In the C complex, $U5^{200}$ remains cross-linked to 5′ splice site sequences and also becomes cross-linked to sequences at the 3′ splice site. In addition, BPS^{56} replaces BPS^{70}, while AG^{75} replaces AG^{100} (Fig. 8).

The novel proteins detected in our study (BPS^{72}, BPS^{70}, BPS^{56}, AG^{75}, and AG^{100}) (Fig. 8) do not cross-link in the hnRNP complex H and are specifically associated with spliceosomal complexes at different stages in spliceosome assembly. In addition, each protein was detected only at a unique location in the pre-mRNA. These results, coupled with the observation that the novel proteins cross-link near or at critical sequence elements, indicate that the factors that we have identified are likely to have important functions in the splicing reaction. In addition, for BPS^{72}, BPS^{56}, AG^{75}, and AG^{100}, mutation of the cross-linking site abolishes cross-linking of the protein. Although we cannot rule out the possibility that the sequence changes affect cross-linking of the proteins, not their binding, the results are at least consistent with the conclusion that these RNA-protein interactions are functionally significant. In previous work, more than 50 proteins were identified as components of purified E, A, B, and C spliceosomal complexes (4, 15), and a similar number of proteins was shown to

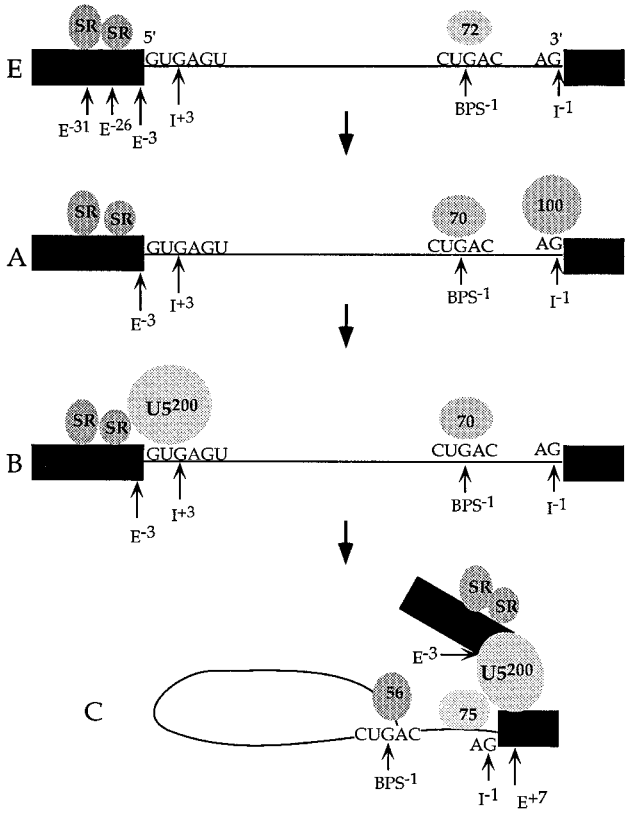


FIG. 8. Summary of proteins that cross-link to the exon, splice sites, and BPS during spliceosome assembly. The positions that were site-specifically ^{32}P labeled in this study are indicated. The cross-linked proteins detected at each site in the spliceosomal complexes (E, A, B, and C) are shown. In this study, we showed that the SR proteins cross-link in exon 1 in the E complex. However, as indicated in the figure, it is likely that the SR proteins remain bound to pre-mRNA throughout spliceosome assembly, as we have detected them cross-linking in the B complex as well (sites were not mapped) (36). Numbers indicate sizes in kilodaltons.

comprise U2, U4/U6, U5, and U4/U5/U6 snRNPs isolated under gentle conditions (1–3, 15, 35). Most of the snRNP proteins are now known to, or are likely to, correspond to the spliceosomal proteins (4, 15, 35). As the novel factors that we have identified were not detected in the previous studies, they must be transiently and/or loosely bound to spliceosomal complexes. Such factors may be detectable only in a sensitive assay such as that used in our study.

A conserved role for U5^{200}. Several strategies have recently been used to detect proteins that cross-link to specific sequences in pre-mRNA. A summary of all of the proteins identified thus far, and the cross-linking reagents used, is presented in Table 1. It is important to note that the precise cross-linking site of a protein is known only for studies that used a photoactivatable nucleotide with a 0-Å cross-linking radius (e.g., 4-thiouridine). When no photoactivatable agent was used (e.g., our study), the cross-linking sites could be mapped only to RNase digestion products containing the ^{32}P-labeled nucleotide. Finally, when benzophenone-derivatized adenosine was used, proteins could be localized to within a 15-Å radius of the derivatized nucleotide (19).

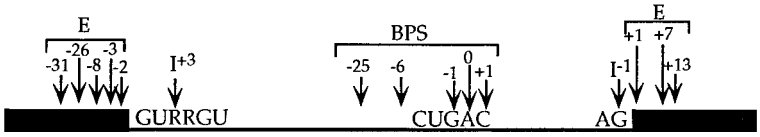
The best-understood cross-linking interaction in the spliceosome, and one that is conserved from *S. cerevisiae* to mammals, is the interaction of U5^{200} first with 5′ splice site sequences in the B complex and later with both 5′ and 3′ splice site sequences in the C complex (Fig. 8). With the caveats noted above for the different cross-linking reagents, U5^{200} can be detected on both exon and intron sequences at both the 5′ and 3′ splice sites (Table 1). At the 5′ splice site, U5^{200} is detected as far upstream as E^{-8} (*S. cerevisiae*) and as far downstream as I^{+3} (mammals). At the 3′ splice site, U5^{200} is detected at the AG dinucleotide in the intron and as far downstream in the exon as E^{+13} (*S. cerevisiae*) (Table 1).

The pre-mRNA–protein interactions of U5^{200} largely parallel the interactions of U5 snRNA with the 5′ and 3′ splice sites; specifically, U5 snRNA can be cross-linked to exon sequences at both splice sites in mammals and interacts genetically with these sequences in yeast cells (10, 25, 26, 34, 42, 44). On the basis of these observations, it has been proposed that U5 snRNA functions to hold exons 1 and 2 together, and possibly align them, for ligation. One possible function of U5^{200} may be in stabilizing the U5 snRNA–pre-mRNA interaction (38, 44). In *S. cerevisiae*, U5^{200} is also thought to play a role in recognition of the pyrimidine tract during catalytic step II of the splicing reaction (41).

Recognition of the branch site. In contrast to the consensus obtained with the U5^{200} cross-linking (references 38, 41, and 44 and this study), significant differences have been observed between different studies on cross-linking at the branch site (Table 1). At least 10 different proteins have been reported to cross-link at the branch site in mammals, and it is not clear which, if any, of these are the same between the different studies. However, differences in the photoactivatable nucleotide used and the locations of these nucleotides in the pre-mRNA may account for many of the differences (Table 1). In our study, two proteins, BPS^{72} and BPS^{56}, are dependent on the branch site adenosine for cross-linking in the E and C complexes, respectively. Further work is needed to determine whether BPS^{70}, which cross-links in the A complex, is also dependent on the BPS for cross-linking. The BPS is not required for assembly of the E complex but does play a role in A-complex assembly (references 8 and 24 and this study). Thus, it is possible that binding of BPS^{72} is important for the E-to-A complex transition. In addition to BPS^{70}, p14, which has been shown to directly contact the branch site adenosine (29), is a good candidate for recognition of this residue during A-complex assembly. The BPS is also required for the catalytic steps of the splicing reaction. The C-complex-specific BPS^{56} and p28 (Table 1) are good candidates for factors involved at this time. p28 cross-linking requires the AG at the 3′ splice site and is therefore more likely to be required for catalytic step II.

Recognition of the AG dinucleotide. The AG dinucleotide plays a role in A-complex assembly, and the requirement for the AG in this step is greater when the pyrimidine tract is weak; in addition, the AG is essential for catalytic step II of the splicing reaction (30, 33). The only protein found thus far that may recognize the AG in the A complex is AG^{100} (Table 1). After A-complex assembly, AG^{100} cross-linking decreases and is replaced by AG^{75} cross-linking in the C complex (Fig. 8; Table 1). Further studies are needed to determine whether either of these proteins is a genuine AG recognition protein or simply interacts with the pre-mRNA in the vicinity of the AG as a result of associations with other proteins. If AG^{100} and AG^{75} do indeed recognize the AG and can account for the roles of these sequence elements in the A and C complexes, respectively, the factors presumably do not recognize the 2-nt sequence alone. Instead, it is likely that the AG is recognized in conjunction with additional sequence elements, such as the adjacent pyrimidine tract. For example, AG^{100} may function by stabilizing the binding of U2AF^{65} to the pyrimidine tract. This would explain why a strong pyrimidine tract partially over-

TABLE 1. Summary of proteins that cross-link to specific sites in pre-mRNA during spliceosome assembly[a]



Sequence element	Protein	Position[b]	Complex	Photoactive agent	Reference
5′ splice site	smD	E^{-2}	A	4-Thiouridine	44
	$U5^{200}$	E^{-2}	B, C	4-Thiouridine	44
	$U5^{200}$	E^{-3}, I^{+3}	B, C	—[c]	This study
	$U5^{200}$	**E^{-1}, E^{-2}, E^{-8}**	**B, C**	**4-Thiouridine**	**38**
3′ splice site	AG^{100}	I^{-1}	A	—[c]	This study
	AG^{75}	I^{-1}	C	—[c]	This study
	$U5^{200}$	I^{-1}, E^{+7}	C	—[c]	This study
	$U5^{200}$	**E^{+1}, E^{+13}**	C	**4-Thiouridine**	**38**
	$U5^{200}$	**I^{-1}**	C	**—[c]**	**41**
Branch site	p80	BPS^{0}	E	Benzophenone-derivatized adenosine	19
	p14, p35, p150	BPS^{0}	A	Benzophenone-derivatized adenosine	19
	$U5^{200}$, p70	BPS^{0}	B/C	Benzophenone-derivatized adenosine	19
	p14,[d] p70	BPS^{0}	A–C	2,6-Diaminopurine	28
	BPS^{72}	BPS^{-1}	E	—[c]	This study
	BPS^{70}	BPS^{-1}	A	—[c]	This study
	BPS^{56}	BPS^{-1}	C	—[c]	This study
	$U2AF^{65}$	BPS^{+1}	—[e]	4-Thiouridine	13
	p28	BPS^{+1}	C	—[c]	13
	SAPs 49, 114, 61, 145, 62, and 155	BPS^{-6} to BPS^{-25} (anchoring site)	A	—[c]	14
Exon	SRp 20, SRp 30	E^{-26} to E^{-31}	E	—[c]	This study

[a] The schematic at the top indicates the sites in pre-mRNA that have been site-specifically labeled for protein cross-linking studies. The mammalian consensus sequences for the splicing elements are shown. All of the proteins so far identified that cross-link at specific sites are listed. Data for *S. cerevisiae* are shown in boldface type.

[b] Site that contains the photoactivatable agent or that was ^{32}P labeled (in cases in which no photoactivatable agent was used).

[c] —, no derivatized nucleotide was used for cross-linking.

[d] p14 was shown to directly contact the adenosine, whereas p70 cross-links to another site within the BPS (29).

[e] $U2AF^{65}$ cross-linking at BPS^{+1} was detected at 4°C in the absence of ATP. None of the known spliceosomal complexes (except for the hnRNP complex H) assemble under these conditions.

comes the requirement for the AG dinucleotide during A-complex assembly (15, 30).

In addition to AG^{75}, we detect $U5^{200}$ cross-linking near or at the AG in the C complex. This protein was also detected at this site in *S. cerevisiae* (Table 1). In contrast to AG^{75} cross-linking, mutation of the AG does not affect $U5^{200}$ cross-linking in *S. cerevisiae* or in mammals.

SR proteins cross-link to exon 1 sequences. Our study revealed that a major site of cross-linking for two members of the SR protein family, SRp 20 and SRp 30, is within exon sequences located about 25 nt upstream of the 5′ splice site. In previous studies, SR proteins have been shown to cross-link to exonic enhancers located downstream of the 3′ splice site; these enhancers function to promote splicing of the upstream intron which contains weak a 5′ or 3′ splice site (see reference 12 for a review). Our data provide the first example of SR proteins interacting with pre-mRNA upstream of the 5′ splice site. How these SR proteins are targeted to this site in exon 1 is not known. The exon 1 site where the SR proteins cross-link does not contain any obvious purine-rich sequences, but non-purine-rich sequences have been shown to bind SR proteins and function as enhancers (36, 39, 40). $U2AF^{65}$ cross-links to a pyrimidine-rich sequence in exon 1 near the SR proteins, but the level of $U2AF^{65}$ cross-linking at this site is significantly lower than that at the 3′ splice site. In any case, SR proteins interact, via protein-protein interactions, with the small subunit of U2AF, $U2AF^{35}$ (43). Thus, it is possible that the SR proteins target U2AF to exon 1 sequences or vice versa. Previous work has shown that exon sequences upstream of the 5′ splice site play a role in recognition of the 5′ splice site (32). This observation, coupled with our finding that SR proteins cross-link upstream of the 5′ splice site, raises the possibility that SR protein-enhancer interactions may occur upstream of the 5′ splice site as well as downstream of the 3′ splice site.

ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We are grateful to Rebecca Feld for excellent technical assistance and to members of the laboratory for useful discussions and comments on the manuscript.

This work was supported by an NIH grant to R.R.

REFERENCES

1. **Bach, M. G., G. Winkelman, and R. Luhrmann.** 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. Proc. Natl. Acad. Sci. USA **86:**6038–6042.
2. **Behrens, S.-E., K. Tyc, B. Kastner, J. Reichelt, and R. Luhrmann.** 1993. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol. Cell. Biol. **13:**307–319.
3. **Behrens, S.-E., and R. Luhrmann.** 1991. Immunoaffinity purification of a U4/U5/U6 tri-snRNP from human cells. Genes Dev. **5:**1439–1452.
4. **Bennett, M., S. Michaud, J. Kingston, and R. Reed.** 1992. Protein components specifically associated with pre-spliceosome and spliceosome complexes. Genes Dev. **6:**1986–2000.
5. **Black, D.** 1995. Finding splice sites within a wilderness of RNA. RNA **1:**763–771.
6. **Brosi, R., K. Groning, S.-E. Behrens, R. Luhrmann, and A. Kramer.** 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relationship of its 60-kD subunit to yeast PRP9. Science **262:**102–105.
7. **Brosi, R., H. P. Hauri, and A. Kramer.** 1993. Separation of splicing factor SF3 into two components and purification of SF3a activity. J. Biol. Chem. **268:**17640–17646.
8. **Champion-Arnaud, P., O. Gozani, L. Palandjian, and R. Reed.** 1995. Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. Mol. Cell. Biol. **15:**5750–5756.
9. **Champion-Arnaud, P., and R. Reed.** 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes Dev. **8:**1974–1983.
10. **Cortes, J. J., E. J. Sontheimer, S. D. Seiwert, and J. A. Steitz.** 1993. Mutations in the conserved loop of U5 snRNA generate use of novel cryptic 5′ splice sites in vivo. EMBO J. **12:**5181–5189.
11. **Eperon, I. C., D. C. Ireland, R. A. Smith, A. Mayeda, and A. R. Krainer.** 1993. Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. EMBO J. **12:**3607–3617.
12. **Fu, X.-D.** 1995. The superfamily of arginine/serine-rich splicing factors. RNA **1:**663–680.
13. **Gaur, R. K., J. Valcarcel, and M. R. Green.** 1995. Sequential recognition of the pre-mRNA branch point by $U2AF^{65}$ and a novel spliceosome-associated 28 kDa protein. RNA **1:**407–417.
14. **Gozani, O., R. Feld, and R. Reed.** 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev. **10:**233–243.
15. **Gozani, O., J. G. Patton, and R. Reed.** 1994. A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J. **13:**3356–3367.
16. **Hodges, B. E., and J. D. Beggs.** 1994. U2 fulfills a commitment. Curr. Biol. **4:**264–267.
17. **Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley.** 1994. Protein-protein interactions and 5′ splice site recognition in mammalian mRNA precursors. Nature (London) **368:** 119–124.
18. **Konarska, M. M., and P. A. Sharp.** 1987. Interactions between snRNPs in formation of spliceosomes. Cell **49:**763–774.
19. **MacMillan, A. M., C. C. Query, C. R. Allerson, S. Chen, G. L. Verdine, and P. A. Sharp.** 1994. Dynamic association of proteins with the pre-mRNA branch region. Genes Dev. **8:**3008–3020.
20. **Madhani, H., and C. Guthrie.** 1994. Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet. **28:**1–26.
21. **Michaud, S., and R. Reed.** 1993. Functional association between the 5′ and 3′ splice sites is established in the earliest pre-spliceosomal complex E. Genes Dev. **7:**1008–1020.
22. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to messenger RNAs by the spliceosome, p. 303–357. *In* R. F. Gesteland and J. F. Atkins (ed.), RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Moore, M. J., and P. A. Sharp.** 1992. Site-specific modification of pre-mRNA: the 2′ hydroxyl groups at the splice sites. Science **256:**992–997.
24. **Nelson, K. K., and M. R. Green.** 1989. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev. **3:**1562–1571.
25. **Newman, A., and C. Norman.** 1991. Mutations in yeast U5 snRNA alter the specificity of 5′ splice site cleavage. Cell **65:**115–123.
26. **Newman, A., and C. Norman.** 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell **68:**743–754.
27. **Newman, A. J.** 1994. Pre-mRNA splicing. Curr. Opin. Genet. Dev. **4:**298–304.
28. **Query, C. C., M. J. Moore, and P. A. Sharp.** 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev. **8:**587–597.
29. **Query, C. C., S. A. Strobel, and P. A. Sharp.** 1996. Three recognition events at the branch site adenosine. EMBO J. **15:**1392–1402.
30. **Reed, R.** 1989. The organization of the 3′ splice-site sequences in mammalian introns. Genes Dev. **3:**2113–2123.
31. **Reed, R.** 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. Curr. Opin. Genet. Dev. **6:**215–220.
32. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46:**681–690.
33. **Smith, C. W. J., E. B. Porro, J. G. Patton, and B. Nadal-Ginard.** 1989. Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. Nature (London) **342:**243–247.
34. **Sontheimer, E. J., and J. A. Steitz.** 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science **262:**1989–1996.
35. **Staknis, D., and R. Reed.** 1994. Direct interactions between pre-mRNA and six U2 small ribonucleoproteins during spliceosome assembly. Mol. Cell. Biol. **14:**2994–3005.
36. **Staknis, D., and R. Reed.** 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol. **14:**7670–7682.
37. **Tacke, R., and J. L. Manley.** 1995. The human splicing factors ASF/SF2 and SC35 possess distinct functionally significant RNA binding specificities. EMBO J. **14:**3540–3551.
38. **Teigelkamp, S., A. J. Newman, and J. D. Beggs.** 1995. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization if exon alignment by U5 snRNA. EMBO J. **14:**2602–2612.
39. **Tian, H., and R. Kole.** 1995. Selection of novel exon recognition elements from a pool of random sequences. Mol. Cell. Biol. **15:**6291–6298.
40. **Tian, M., and T. Maniatis.** 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell **74:**105–114.
41. **Umen, J. G., and C. Guthrie.** 1995. A novel role for a U5 snRNP protein in 3′ splice site selection. Genes Dev. **9:**855–868.
42. **Wassarman, D. A., and J. A. Steitz.** 1992. Interactions of small nuclear RNAs with precursor messenger RNA during in vitro splicing. Science **257:**1918–1925.
43. **Wu, J. Y., and T. Maniatis.** 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75:**1061–1070.
44. **Wyatt, J. R., E. J. Sontheimer, and J. A. Steitz.** 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.
45. **Zamore, P. D., and M. R. Green.** 1989. Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc. Natl. Acad. Sci. USA **86:**9243–9247.
46. **Zuo, P., and J. L. Manley.** 1994. The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites. Proc. Natl. Acad. Sci. USA **91:**3363–3367.