Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals

Maria Dolores Chiara¹, Leon Palandjian, Rebecca Feld Kramer and Robin Reed²

Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

¹Present address: Departamento de Fisiología Médica y Biofísica, Facultad de Medicina, Universidad de Sevilla, E-41009, Sevilla, Spain

²Corresponding author

M.D.Chiara and L.Palandjian contributed equally to this work

The first AG dinucleotide downstream from the branchpoint sequence (BPS) is chosen as the 3' splice site during catalytic step II of the splicing reaction. The mechanism and factors involved in selection of this AG are not known. Early in mammalian spliceosome assembly, U2AF⁶⁵ binds to the pyrimidine tract between the BPS and AG. Here we show that U2AF⁶⁵ crosslinking is replaced by crosslinking of three proteins of 110, 116 and 220 kDa prior to catalytic step II, and we provide evidence that all three proteins are components of U5 snRNP. These proteins interact with pre-mRNA in the region spanning from immediately downstream of U2 snRNP's binding site at the BPS to just beyond the 3' splice site. We also demonstrate that there are strict constraints on both the sequence and the distance between the BPS and AG for catalytic step II. Together, these observations suggest that U5 snRNP is positioned on the 3' splice site by an interaction (direct or indirect) with U2 snRNP bound at the BPS and by a direct interaction with the pyrimidine tract. The functional AG for catalytic step II may be specified, in turn, by its location with respect to the U5 snRNP binding site.
Keywords: 3' splice site/spliceosome/splicing/U5 snRNP

Introduction

Pre-mRNA splicing takes place via assembly of a series of highly dynamic spliceosomal complexes followed by two successive transesterification reactions. The spliceosomal complexes assemble on pre-mRNA in the order E→A→B→C, with the catalytic steps of splicing occurring in the C complex. In the first transesterification reaction, the 5' splice site is cleaved to generate the splicing intermediates (exon 1 and lariat-exon 2), and in the second reaction, the 3' splice site is cleaved to generate the spliced products (lariat intron and spliced mRNA). The high fidelity of splicing is achieved through networks of interactions involving more than 50 distinct spliceosomal proteins and five small nuclear RNAs (U1, U2, U4, U5 and U6) (for reviews, see Adams *et al.*, 1996; Kramer, 1996; Reed, 1996).

U2, U5 and U6 snRNAs each interact with specific

sites in pre-mRNA very near or at the catalytic center of the spliceosome, and these interactions are detected immediately prior to the two steps of splicing (Parker *et al.*, 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989; Newman and Norman, 1991, 1992; Sawa and Abelson, 1992; Sawa and Shimura, 1992; Wassarman and Steitz, 1992; Wyatt *et al.*, 1992; Cortes *et al.*, 1993; Kandels-Lewis and Seraphin, 1993; Lesser and Guthrie, 1993; Sontheimer and Steitz, 1993; Newman *et al.*, 1995; O'Keefe *et al.*, 1996). One or more of these snRNAs are believed to be the catalytic moieties of the spliceosome (for reviews, see Madhani and Guthrie, 1994; Newman, 1994; Nilsen, 1994; Kramer, 1996). In contrast to other examples of RNA-mediated catalysis, proteins play crucial roles in splicing. However, the technical difficulty of pinpointing specific binding sites of proteins on RNA has hampered progress in identifying RNA–protein interactions at the catalytic center of the spliceosome.

One step for which little information is available is selection of the 3' splice site during catalytic step II. In both yeast and metazoans, the 3' splice site consists of YAG (Y = pyrimidine), and the pre-mRNA is cleaved after the G residue during catalytic step II. The YAG is preceded by a pyrimidine tract, with this element being less conserved in yeast. In most introns, the first AG downstream of the branchpoint sequence (BPS) serves as the 3' splice site, and this AG is typically located within 20–40 nt of the BPS. Although the mechanism and factors involved in selecting the correct AG are not understood, considerable progress has been made in the yeast system (for review, see Umen and Guthrie, 1995a). Analysis of mutant pre-mRNAs has revealed that a uridine-rich tract adjacent to the AG enhances the efficiency of step II (Patterson and Guthrie, 1991). In addition, although there is a clear preference for branch site-proximal AGs, a downstream AG can compete with an upstream AG when preceded by a uridine-tract (Patterson and Guthrie, 1991). Finally, during preparation of our manuscript, Luukkonen and Seraphin (1997) reported that efficient selection of the 3' splice site depends on an optimal distance between the BPS and AG (13–22 nt).

Considerable progress has also been made in identifying proteins that play a role in 3' splice-site recognition in yeast. The best characterized of these are PRPs 8 and 16 and Slu7 (for review, see Umen and Guthrie, 1995a). PRP8 is an integral U5 snRNP component (Lossky *et al.*, 1987; Whittaker *et al.*, 1990) and the others interact genetically with U5 snRNP (Frank *et al.*, 1992; Umen and Guthrie, 1995c). *In vitro* studies indicate that PRP16 crosslinks to the 3' splice site, followed by crosslinking of Slu7 and PRP8 (Teigelkamp *et al.*, 1995; Umen and Guthrie, 1995b,c). Although the precise locations of these proteins on the 3' splice site are not known, genetic studies suggest a direct or indirect role for PRP8 in

recognition of the U-rich tract (Umen and Guthrie, 1995b) and YAG (Umen and Guthrie, 1995b, 1996). In contrast, SLU7 functions in use of 3' splice sites located further (more than ~10 nt) from the branch site (Frank and Guthrie, 1992; Byrs and Schwer, 1996).

PRP8 (Teigelkamp *et al.*, 1995) and two snRNAs, U2 and U5 (Newman *et al.*, 1995), also crosslink to exon sequences downstream of the 3' splice site prior to catalytic step II. In addition, U5 snRNA and PRP8 crosslink to exon sequences upstream of the 5' splice site at the same time (Newman *et al.*, 1995; Teigelkamp *et al.*, 1995). Essentially the same interactions of the two exons with U5 snRNA (Sontheimer and Steitz, 1993) and the mammalian homologue of PRP8 (U5²²⁰) (Chiara *et al.*, 1996) occur prior to step II in mammals. On the basis of all of the yeast and mammalian data, a model for one stage in catalytic step II has been proposed (Sontheimer and Steitz, 1993; Newman *et al.*, 1995; Teigelkamp *et al.*, 1995; O'Keefe *et al.*, 1996). In this model, U5 snRNA, aided by PRP8/U5²²⁰, 'grips' and possibly aligns the exons for ligation. Consistent with this model, O'Keefe and co-workers (1996) have functional data in yeast that indicate a role for U5 snRNA in holding the exons together for catalytic step II. However, exon sequences near the 5' and 3' splice sites are not well conserved in either yeast or metazoans. Thus, the U5 snRNP–exon interactions cannot be a major determinant of 3' splice-site selection.

In contrast to the progress made in understanding recognition of the 3' splice site in yeast, very little is known in metazoans. As in yeast, there is evidence that both the sequence and the distance between the BPS and 3' splice site play a role in step II (Reed, 1989; Smith *et al.*, 1993). However, no systematic analysis of these parameters has been carried out. In addition, it is not known what factors mediate these effects. The factors so far implicated in 3' splice-site recognition for step II in mammals include PSF (Gozani *et al.*, 1994) and AG⁷⁵ (Chiara *et al.*, 1996). PSF binds to pyrimidine-rich sequences (Patton *et al.*, 1993), but this protein has not been shown to interact at the 3' splice site in functional spliceosomal complexes. AG⁷⁵ crosslinks to pre-mRNA very close to or at the AG prior to step II and thus is the only good candidate identified in mammals for a 3' splice-site recognition factor (Chiara *et al.*, 1996). Other factors that play a role in step II include U5²⁰⁰ (Lauber *et al.*, 1996) and the human homologue of the yeast step II protein PRP18 (Horowitz and Krainer, 1997), but no data indicate a specific role for these proteins in 3' splice-site selection.

The scarcity of information on recognition of the 3' splice site for step II in metazoans prompted us to identify all of the proteins that interact in the vicinity of the 3' splice site and to investigate the possible role of these interactions in selection of the 3' splice site for step II. This analysis revealed that U2AF⁶⁵, which crosslinks to the pyrimidine tract early in spliceosome assembly, is replaced by three distinct proteins (110, 116 and 220 kDa) late in spliceosome assembly, and prior to catalytic step II. Significantly, our data indicate that all three proteins are components of U5 snRNP, revealing a key role for this snRNP in 3' splice-site recognition for step II in mammals. A systematic analysis of mutant pre-mRNAs showed that the presence of pyrimidines and the distance between the BPS and AG are critical determinants of

catalytic step II efficiency. These sequence/distance constraints, correlated with the RNA–protein interactions that span from the BPS to the 3' splice site, suggest that U2–U5 snRNP interactions (direct or indirect) play a role in selecting the 3' splice site for catalytic step II. Recent studies support the possibility that a similar network of interactions is involved in 3' splice-site selection in yeast (Byrs and Schwer, 1996; Lauber *et al.*, 1996; Xu *et al.*, 1996; Luukkonen and Seraphin, 1997).

Results

Recognition of the 3' splice site for catalytic step II

The pyrimidine tract at the 3' splice site is first recognized in the spliceosomal complex E where it is bound by the essential splicing factor U2AF⁶⁵. To identify factors that recognize the pyrimidine tract at subsequent stages in spliceosome assembly, we used a site-specific labeling/UV crosslinking strategy. In this strategy, pre-mRNA containing a single ³²P-labeled guanosine (Moore and Sharp, 1992) is assembled into complexes at different stages of spliceosome assembly. These complexes are then isolated by gel filtration, UV-crosslinked and treated with RNase A; the proteins crosslinked to the labeled RNase digestion product are detected on SDS or 2D gels (Gozani *et al.*, 1996). Due to the combined technical difficulties of obtaining large quantities of the site-specifically labeled pre-mRNAs and isolating each complex by gel filtration, we opted to limit our study to comparing well-characterized stages of spliceosome assembly (i.e. the E, A/B and C complexes).

Catalytic steps I and II take place in the C complex. Maximal levels of the step I products (exon 1 and lariat-exon 2) are detected with wild-type AdML pre-mRNA (see schematic, Figure 1A) when incubated under splicing conditions for 30 min (Figure 1B). This time point was therefore used as our C complex preparation. The E, A/B and H complexes were analyzed for comparison (A/B is primarily A complex contaminated with a small amount of B complex; see Materials and methods).

Initially, we examined pre-mRNA ³²P-labeled at a guanosine within the pyrimidine tract 6 nt upstream from the 3' splice junction [Figure 1A, I⁻⁶ (–6 position of the intron)]. After spliceosomal complexes were isolated by gel filtration and UV crosslinked, total RNA or protein was prepared (Figure 1C and D). As expected, only pre-mRNA was detected in gel filtration-isolated A/B complex, whereas pre-mRNA and the lariat-exon were detected in the C complex (Figure 1C). Analysis of the proteins revealed that U2AF⁶⁵ crosslinks with similar efficiency in the E through C complexes (Figure 1D, lanes 2–4). A novel 140 kDa protein first crosslinks in the A/B complex and is also detected in the C complex (Figure 1D, lanes 3 and 4). Proteins of 110 and 116 kDa crosslink in the C complex, but do not crosslink in any of the other complexes (Figure 1D, compare lane 4 with lanes 2, 3 and 5). A 220 kDa protein can also be detected in the C complex on long exposures (data not shown). Finally, a number of proteins detected in the A/B and C complexes are due to contamination with the hnRNP complex H (Figure 1D, unlabeled bands, compare lane 5 with lanes 2–4; note that the H complex lane was underloaded in this experiment).

It was initially puzzling that several high molecular

Fig. 1. Proteins crosslinked to the pyrimidine tract at different steps in spliceosome assembly. **(A)** Schematic of WT AdML pre-mRNA. The nucleotide sequence between the branchpoint sequence (BPS) and the 3' splice site is shown. The BPS is underlined; the branchpoint adenosine and 3' splice-site AG dinucleotide are enlarged and in bold type. I⁻⁶ indicates the nucleotide that was changed from C to G and ³²P-labeled (see Materials and methods). **(B)** Uniformly labeled WT pre-mRNA was incubated in splicing reactions for the times indicated and then total RNA was fractionated on a 15% denaturing polyacrylamide gel. **(C)** Total RNA from gel filtration-isolated A/B or C complexes, which were assembled on WT pre-mRNA labeled at I⁻⁶, was fractionated on a 15% denaturing polyacrylamide gel. **(D)** Crosslinked proteins detected in the E, A/B, C or H complexes assembled on WT pre-mRNA labeled at I⁻⁶. Proteins were fractionated on a 9% SDS-polyacrylamide gel. U2AF⁶⁵ was *in vitro*-translated from a cDNA and fractionated as a marker in lane 1. Crosslinked U2AF⁶⁵ was also identified by immunoprecipitation using U2AF⁶⁵ antibodies (data not shown). The bands corresponding to U2AF⁶⁵, 140, 116 and 110 kDa proteins are indicated. The unlabeled bands are hnRNP proteins. The molecular weight markers indicated with bars on the left are 206, 117, 89 and 47 kDa.

weight proteins crosslink in the C complex without a corresponding decrease in the proteins that crosslink early in spliceosome assembly (U2AF⁶⁵ and the 140 kDa protein) (Figure 1D). However, only 25% of the pre-mRNA is converted to lariat-exon in our C complex preparation (Figure 1C) which is heavily contaminated with the earlier spliceosomal complexes (A and B complexes) (data not shown). Due to the rapid kinetics of catalytic step II on wild-type pre-mRNA (see Figure 1B, 30 min and 45 min time points), it is not possible to obtain a C complex highly enriched in the splicing intermediates. Thus, we could not distinguish whether the crosslinking of the 110, 116, and 220 kDa proteins occurs in the B or in the C complex. This issue prompted us to analyze GG-GG pre-mRNA in which the AG at the 3' splice site is substituted with a GG (see schematic, Figure 2A). A cryptic AG located 6 nt further downstream is also substituted with a GG (see Figure 6A for the sequence in this region). As shown in Figure 2B, these substitutions block catalytic step II and result in accumulation of the C complex containing the splicing intermediates.

To identify proteins that crosslink to the 3' splice site of GG-GG pre-mRNA, we ³²P-labeled at the -6 site (Figure 2A). In the gel filtration-isolated C complex assembled on this pre-mRNA, >60% of the pre-mRNA is converted to lariat-exon (Figure 2C, left panel; the faint band below the lariat is a breakdown product). Strikingly, the 220, 116 and 110 kDa proteins crosslink much more efficiently on GG-GG than on wild-type (WT) pre-mRNA (relative to U2AF⁶⁵, compare Figure 2C, right panel, lane 3 with Figure 1D, lane 4). Moreover, on GG-GG pre-mRNA, the levels of U2AF⁶⁵ crosslinking in the C

complex are significantly lower than in the A/B complex (Figure 2C, right, compare lanes 2 and 3). From analyzing a large number of independent preparations of the C complex, we find that there is a strong inverse correlation between the levels of U2AF⁶⁵ crosslinking and the levels of lariat-exon (data not shown). Our data do not indicate whether U2AF⁶⁵ dissociates entirely from the C complex or simply undergoes a conformational change that results in loss of crosslinking. In either case, we conclude that a major rearrangement of RNA-protein interactions occurs on the 3' splice site prior to catalytic step II of the splicing reaction. The levels of the 140 kDa protein are also decreased in the C complex relative to the A/B complex (Figure 2C, right, lanes 2 and 3), but the levels of this protein vary in different preparations of the complexes (data not shown and see below).

Two-dimensional gel analysis confirmed that the high molecular weight proteins detected on GG-GG pre-mRNA are the same as those detected on WT pre-mRNA (data not shown). Thus, we conclude that on both WT and GG-GG pre-mRNA, crosslinking of U2AF⁶⁵ (and possibly the 140 kDa protein) early in spliceosome assembly is replaced by crosslinking of 110, 116 and 220 kDa proteins later in spliceosome assembly (Figures 1 and 2). To determine whether there is a temporal binding order for these three proteins, we compared the crosslinked proteins at the I⁻⁶ site in the B and C complexes assembled on GG-GG pre-mRNA (Figure 2D). The 116 kDa protein was detected in both the B and C complexes whereas the 110 and 220 kDa proteins were detected only in the C complex (Figure 2D, lanes 2 and 3). We conclude that the 116 kDa

Fig. 2. Crosslinked proteins detected on the pyrimidine tract in the C complex. (A) Schematic of GG-GG pre-mRNA. The GG at the 3' splice site and the ^{32}P labeling site (I^{-6}) are shown. (B) Uniformly labeled GG-GG pre-mRNA was incubated in splicing reactions for the times indicated. (C) Right panel: crosslinked proteins in H, A/B and C complexes assembled on GG-GG pre-mRNA labeled at I^{-6} (the A/B complex was assembled by incubation for 7.5 min under splicing conditions). Left panel: total RNA from the gel filtration-isolated complexes shown in the right panel. (D) Crosslinked proteins in the H, B and C complexes (the B complex was assembled by incubation for 15 min under splicing conditions). The crosslinked bands specific to the complexes are labeled on the right, and molecular weight markers (kDa) and gel origin (ori) are shown on the left. The unlabeled proteins are hnRNP proteins.

protein crosslinks to the 3' splice site prior to the 110 and 220 kDa proteins.

Several observations provide evidence that the 110, 116 and 220 kDa proteins interact with the 3' splice site prior to step II. First, in many independent preparations of the C complex, in which the conversion to splicing intermediates varies, we found a direct correlation between the levels of the 110 and 220 kDa crosslinked proteins and the levels of splicing intermediates. Second, high levels of all three crosslinked proteins are still detected at very long time points (80 and 90 min, data not shown), in contrast to U2AF⁶⁵ crosslinking which diminishes over time. Finally, we have not detected any other proteins that replace the 110, 116 and 220 kDa proteins at later time points (data not shown).

Identification of the crosslinked proteins

To determine whether any of the newly identified crosslinked proteins corresponds to known spliceosomal proteins (Bennett *et al.*, 1992; Gozani *et al.*, 1994), we used gel filtration followed by biotin-avidin affinity selection to isolate the C complex assembled on pre-mRNA site-specifically labeled at I^{-6} . Analysis of the crosslinked proteins on 2D gels (data not shown) revealed

that the 220 and 116 kDa crosslinked proteins precisely co-migrate with silver-stained U5 snRNP proteins, U5²²⁰ and U5¹¹⁶ (Bach *et al.*, 1989), present in the C complex (Gozani *et al.*, 1994). The 110 kDa crosslinked protein was not detected in the gel filtration/affinity-selected complex, indicating that it dissociates during the affinity purification. To confirm that the 220 and 116 kDa proteins are U5 snRNP proteins, we mixed crosslinked C complex (isolated by gel filtration alone) with purified 25S U5 snRNP and compared the silver-stained and crosslinked proteins on a 2D gel (Figure 3A). This analysis revealed that all three crosslinked proteins co-migrate with U5 snRNP proteins (Figure 3A, compare panels II and III). U5²²⁰, U5¹¹⁶ and U5¹¹⁰ co-migrate with crosslinked 220, 116 and 110 kDa proteins, respectively. We note that although crosslinked RNase A digestion products can significantly alter the mobilities of low molecular weight proteins (Chiara *et al.*, 1996), we have not observed an alteration in the migration of high molecular weight proteins (Staknis and Reed, 1994; Gozani *et al.*, 1996).

PSF was previously speculated to recognize the pyrimidine tract for step II (Gozani *et al.*, 1994). However, we find that PSF does not co-migrate on silver-stained 2D gels with any of our crosslinked proteins (data not shown).

Fig. 3. Evidence that the 220, 116 and 110 crosslinked proteins are U5 snRNP components. (A) Comparison of U5 snRNP proteins and crosslinked C complex proteins by 2D gel electrophoresis. I, U5 snRNP proteins. II, III, U5 snRNP proteins mixed with crosslinked C complex assembled on GG-GG pre-mRNA labeled at I⁻⁶. Panels I and II show silver-stained gels; panel III shows a phosphorimager pattern of the gel in panel II. The U5 snRNP-specific proteins (Bach *et al.*, 1989) are indicated. (B) Immunoprecipitation of crosslinked proteins with antibodies to U5 snRNP proteins. The indicated complexes were assembled on GG-GG pre-mRNA labeled at I⁻⁶. After treatment with RNase A and UV crosslinking, gel-filtration fractions from the C complex peak were immunoprecipitated with the indicated antibodies under denaturing or non-denaturing conditions (see Materials and methods). Crosslinked proteins detected specifically in the C complex are labeled on the left.

As the evidence that PSF is a pyrimidine tract-binding protein was based solely on its interactions with naked RNA (Patton *et al.*, 1993), it is quite likely that PSF binds at some other location on the pre-mRNA.

To obtain further support for the identity of the crosslinked 110, 116 and 220 kDa proteins, we carried out immunoprecipitations using polyclonal antibodies to U5²²⁰ (generous gift from M. Moore), and a monoclonal antibody, mAb386, which recognizes the U1 snRNP component U1 70K and crossreacts with U5¹¹⁰ (generous gift from R. Luhrmann; Behrens and Luhrmann, 1991). The U5²²⁰ antibody immunoprecipitates well only under

denaturing conditions (G. Moreau and M. Moore, personal communication), whereas mAb386 only cross-reacts with U5¹¹⁰ under non-denaturing conditions (Behrens and Luhrmann, 1991). Significantly, the U5²²⁰ antibody specifically immunoprecipitates the crosslinked 220 kDa protein under denaturing conditions (Figure 3B, lane 5) from the total crosslinked proteins in the C complex (lane 2). With mAb386, we observed immunoprecipitation of the 220, 140, 116 and 110 kDa proteins and U2AF⁶⁵ (Figure 3B, lane 4); all of these proteins were selectively immunoprecipitated relative to the low molecular weight hnRNP protein present in the total C complex (Figure 3B,

Fig. 4. U5 snRNP protein-pre-mRNA interactions span from just downstream of the BPS to just downstream of the 3' splice site. (A) Schematic of GG-GG pre-mRNA showing G residues that were ^{32}P -labeled. (To synthesize transcripts labeled at the I⁻¹³ and I⁻⁶ sites, it was necessary to change the sequences at these positions from UU to GC and from C to G, respectively; see Materials and methods.) (B) Crosslinked proteins detected in the C complex when pre-mRNA was labeled at the indicated sites. Crosslinked complexes were treated with RNase A or T1 as designated. Specific crosslinked proteins are labeled, and the unlabeled bands are hnRNP proteins. The band in lane 7, labeled with *, corresponds either to a modified U5¹¹⁰ or to a novel 110 kDa protein (see text). (C) RNase protection. Naked pre-mRNA or gel filtration-isolated C or H complexes were treated with the indicated RNases under the same conditions used for the crosslinked complexes (see Materials and methods). Total RNA was prepared and fractionated on a 20% denaturing polyacrylamide gel. The sizes of the digestion products are indicated. The protected band is indicated by the **.

lane 2). We also carried out an immunoprecipitation of the C complex using another monoclonal antibody to U1 70K (Billings *et al.*, 1982), which recognizes an epitope similar to that of mAb386 (Spritz *et al.*, 1987; Behrens and Luhrmann, 1991). mAb U1 70K (Figure 3B, lane 8) immunoprecipitated the same set of proteins as mAb386 from the total C complex proteins (lane 7). These proteins are not immunoprecipitated by control antibodies (lane 9 and data not shown).

On the basis of the co-migration and immunoprecipitation data, we conclude that the 220 kDa crosslinked protein is U5²²⁰. The 116 kDa crosslinked protein is likely to be U5¹¹⁶ because these proteins co-migrate precisely on 2D gels and on different percentage SDS gels, and both proteins are present in C complex purified by gel filtration/affinity selection. As the identity of the 116 kDa protein could not be confirmed by other methods, the possibility that there may be two different 116 kDa proteins of identical 2D gel mobility that crosslink in affinity-selected C complex cannot be formally excluded. In the case of the 110 kDa crosslinked protein, we find that it co-migrates with U5¹¹⁰ on a 2D gel. Further evidence that these proteins are the same is that, while U5¹¹⁰ is a snRNP component and the 110 kDa protein crosslinks in the C complex, neither protein is detected in affinity-selected C complex (data not shown, Gozani *et al.*, 1994). The immunoprecipitation studies with mAb386 and the U1 70K mAb are also consistent with the notion that U5¹¹⁰ and the 110 kDa crosslinked proteins are the same. Both of the antibodies share an epitope that is characterized by repeating arginine/aspartate (R/D) residues (Spritz

et al., 1987; Behrens and Luhrmann, 1991). Thus, one interpretation of the immunoprecipitation data is that these antibodies recognize U5¹¹⁰ directly and co-immunoprecipitate U5²²⁰ and the 116 kDa protein. Immunoprecipitation of U2AF⁶⁵ is most likely direct because it has previously been shown that mAb U1 70K and another antibody that recognizes RD/RE epitopes immunoprecipitates U2AF⁶⁵ (Neugebauer *et al.*, 1995; Staknis and Reed, 1995). However, given the complexity of the immunoprecipitation data with mAb386 and mAb U1 70K, we cannot rule out other interpretations of the data. We were unable to identify the 140 kDa crosslinked protein as a known spliceosomal protein, and thus have designated it py¹⁴⁰.

Arrangement of proteins on the 3' splice site

To characterize all of the RNA-protein interactions in the vicinity of the 3' splice site, we ^{32}P -labeled GG-GG pre-mRNA at different locations and analyzed the crosslinking patterns after digestion with RNase A (cleaves after pyrimidines) or RNase T1 (cleaves after guanines). All proteins were identified by comparing their migration on SDS and/or 2D gels to the set of crosslinked proteins detected at I⁻⁶ (Figure 4 and data not shown). We found that this same set of proteins is also detected on the 19 nt RNase T1 fragment (underlined, Figure 4A) when pre-mRNA is labeled at I⁻¹ (Figure 4B, compare lanes 1 and 4). In addition, we detect a new band that migrates between the 116 kDa protein and py¹⁴⁰ (lane 1); it is not clear whether this new band is a sub-population of the crosslinked 116 kDa protein that shifts up due to the large

crosslinked T1 fragment or a new protein that is not seen on any of the RNase A fragments (see below).

The same crosslinking pattern observed on the 19 nt T1 fragment is observed on the 13 nt T1 fragment (underlined twice, Figure 4A) when pre-mRNA is labeled at I⁻⁶ (Figure 4B, compare lanes 1 and 3). Thus, all of the crosslinked proteins are present on the RNase T1 fragment spanning from I⁻⁶ to I⁻¹⁹. Consistent with this conclusion, the 110, 116 and 220 kDa proteins are barely detected at I⁻¹ after RNase A digestion (Figure 4B, lane 5). Interestingly, however, U5²²⁰ is detected strongly further downstream at E⁺⁶ after RNase A digestion (Figure 4B, compare lanes 6 and 7; little crosslinking is detected at E⁺⁶ after RNase T1 digestion, most likely because of the high G content in this region, compare lanes 1 and 2). A 110 kDa band also crosslinks at the E⁺⁶ site (Figure 4B, lane 7, indicated by *); on 2D gels, this protein does not co-migrate with U5¹¹⁰ (data not shown), and thus appears to be a novel protein.

In the region spanning from I⁻⁶ to I⁻¹⁹, the 116 kDa protein crosslinks to pre-mRNA upstream of the 110 kDa protein and U5²²⁰. This conclusion is based on the finding that the ratio of the 116 kDa protein to both U5²²⁰ and the 110 kDa protein is much greater at I⁻¹³ than at I⁻⁶ (Figure 4B, compare lanes 8 and 9). We were unable to establish the relative positions of U5²²⁰ and the 110 kDa protein in the I⁻⁶ to I⁻¹⁹ region. It is possible that these two proteins bind to the same site in different C complex subpopulations (which may be present due to heterogeneity in the preparation). Alternatively, these proteins may bind so closely to each other that they protect the RNA from RNase digestion, thus making it difficult to determine their relative locations (see below). Consistent with the latter possibility, the RNA in the region from I⁻⁶ to I⁻¹⁹ appears to be partially protected from RNase because the RNase T1 and RNase A digestion patterns at the I⁻⁶ site are so similar, despite the presence of many potential RNase A cleavage sites (see Figure 4A; also compare lanes 3 and 4 in Figure 4B).

To determine directly whether the pre-mRNA in the vicinity of the 3' splice site is protected from RNase digestion in the C complex, we analyzed total pre-mRNA after treatment of the I⁻⁶ C complex with RNase T1 or A (Figure 4C). After T1 digestion, the expected 13 nt fragment was detected (Figure 4C, lane 3); this 13 nt fragment co-migrated with the T1 digestion product generated from naked pre-mRNA labeled at I⁻⁶ (Figure 4C, lane 5). After RNase A digestion of the C complex, >40% of the RNA was detected in bands (designated with **) that are significantly larger than the dinucleotide detected when naked pre-mRNA is digested with RNase A (Figure 4C, compare lanes 4 and 6). We conclude that the pre-mRNA in the vicinity of I⁻⁶ is partially protected from RNase A digestion in the C complex. This region is also partially protected in the H complex (Figure 4C, lane 2), presumably due to the multiple hnRNP proteins that crosslink in this complex (e.g. see Figure 2C and D). The protection of the pre-mRNA in the C complex provides the most likely explanation for the detection of a similar set of proteins at the I⁻⁶ site after RNase T1 or RNase A digestion.

On the basis of the crosslinking data and the protection analysis, we conclude that at least three proteins of 110,

Fig. 5. Evidence that steric hindrance by U5¹¹⁶ blocks use of an AG located too close to the branch site. (A) Schematics of WT, 15AG and 11AG pre-mRNAs. The sequences and number of nucleotides between the branchpoint adenosine (enlarged) and the duplicated AGs are shown. The AGs that function as 3' splice sites are enlarged and bolded. (B) Splicing of WT, 15AG and 11AG pre-mRNAs. Each pre-mRNA was incubated under splicing conditions for 60 min, and total RNA was fractionated on an 8% denaturing polyacrylamide gel. (C) Crosslinked proteins detected in the C complex assembled on ³²P site-specifically labeled 11AG pre-mRNA. ³²P-labeling sites (I⁻⁸ and I⁻¹) are shown with arrows in (A). Specific crosslinked proteins and the gel origin (ori) are indicated.

116 and 220 kDa crosslink to the pyrimidine tract in the C complex, spanning the distance from just downstream of the BPS to just downstream of the 3' splice site. It is not possible to determine whether all of these proteins contact the pre-mRNA at the same time or sequentially. In any case, our data suggest that the 116 kDa protein binds closest to the BPS, and U5²²⁰ contacts both sides of the 3' splice site. The 116 kDa protein is the closest crosslinked protein to the U2 snRNP protein SAP 155 which is the only protein detected at I⁻¹⁹ in the C complex (O.Gozani and R.Reed, unpublished observations). Finally, we have previously detected U5²²⁰ crosslinking at the E⁺⁷ site on wild-type pre-mRNA (Chiara *et al.*, 1996). Thus, the positioning of U5²²⁰ on both sides of the 3' splice site occurs independently of the AG dinucleotide.

Selection of the AG dinucleotide

The first AG downstream of the BPS functions as the 3' splice site during catalytic step II, and this AG is usually located between 18–40 nt from the BPS. To investigate the mechanism for selection of this AG and the function of the crosslinked proteins in this process, we examined pre-mRNAs containing AGs located in different sequence contexts and at varying distances from the BPS. We first constructed pre-mRNAs containing duplicated AGs in which the upstream AG was moved progressively closer to the BPS (Figure 5A). When the AG is located at the wild-type position, 23 nt from the BPS, splicing occurs

Fig. 6. Effect of increasing the BPS-to-AG distance. (A) Schematics of pre-mRNAs. Only the 3' portion of each pre-mRNA is shown. The sequences of the regions that differ among the pre-mRNAs are indicated. (B) WT, GG-AG, GG----AG, and GG-GG were incubated for 40 min under splicing conditions. The ratio of spliced mRNA to exon 1 is indicated below the gel (quantitated by phosphorimager). (C) WT and rat α -tropomyosin pre-mRNA were incubated under splicing conditions for the times indicated. Total RNA was fractionated on a 15% denaturing polyacrylamide gel (B and C).

at the upstream AG (Figure 5B, WT, lanes 1 and 4). The upstream AG is also selected when it is located 15 nt from the BPS (15AG, Figure 5A; see also Figure 5B, lane 2). In contrast, the upstream AG is completely skipped when it is moved to a position 11 nt away from the BPS, and splicing occurs at the downstream AG (Figure 5B, 11AG, lane 3). Thus, the 4 nt difference between the otherwise isogenic 11AG and 15AG pre-mRNAs is sufficient to completely switch the AG selected.

To understand the basis for skipping the upstream AG in 11AG pre-mRNA, we compared the proteins that crosslink at the upstream AG (I^{-8} , Figure 5A) to those that crosslink at the downstream functional AG (I^{-1} , Figure 5A). Significantly, we found that AG⁷⁵, a protein previously shown to crosslink to functional AGs in wild-type pre-mRNA (Chiara *et al.*, 1996), crosslinks strongly at the I^{-1} site, but not at the I^{-8} site (Figure 5C, lane I^{-1}). Instead, crosslinking of the 116 kDa protein is detected at I^{-8} (Figure 5C, I^{-8} ; note that py¹⁴⁰ and U2AF⁶⁵ are also detected, most likely due to contamination of the C complex with A and B complexes; see above). Significantly, the I^{-8} site in 11AG is in the same position relative to the BPS as the I^{-1} site, Figure 4B, lane 8). These data suggest that AG dinucleotides located too close to

the BPS are not used due to steric hindrance from the 116 kDa protein.

We next investigated pre-mRNAs in which the AG was moved progressively further from the BPS (Figure 6A). As observed above with WT pre-mRNA, the upstream AG, located 23 nt from the BPS, is used exclusively (Figure 6B, WT). When this AG is mutated to GG, the AG located 6 nt downstream is efficiently used (Figure 6B, GG-AG). However, when the AG is moved an additional 6 nt downstream, there is a significant decrease (3- to 4-fold) in the efficiency of catalytic step II (Figure 6B, GG----AG). The only difference between GG-AG and GG----AG pre-mRNAs is the 6 nt between the GG and the AG. The observation that catalytic step II occurs efficiently over such a narrow range of BPS-to-AG distance (4 nt shorter or 6 nt longer than WT has a major impact on step II efficiency), is a striking observation in light of previous work. Most notably, in rat α -tropomyosin intron 2, the first AG is located as far as 180 nt downstream from the BPS, yet it functions as the 3' splice site (Smith *et al.*, 1989). To investigate the apparent inconsistency between this observation and our results, we compared splicing time courses of WT AdML and α -tropomyosin pre-mRNAs (Figure 6C). Significantly, this comparison revealed that catalytic step II has barely begun by 1 h

with α -tropomyosin whereas significant levels of spliced mRNA have already accumulated by 30 min with WT AdML (Figure 6C). By the 2-h time point, high levels of intermediates remain with α -tropomyosin but not with WT AdML, indicating the relative inefficiency of catalytic step II with α -tropomyosin pre-mRNA.

long as 2 h, whereas catalytic step II is nearly complete by 1 h with WT pre-mRNA (Figure 7B, compare WT with Ran mutants).

The best of the pyrimidine insertion mutants for catalytic step II is pyJ, which has several uridine tracts near the AG (Figure 7C). However, comparing WT and pyJ at

several time points over a longer splicing time course revealed that even pyJ does not undergo catalytic step II as efficiently as WT (Figure 7B, compare WT with pyJ, and data not shown). This observation suggests that the BPS-to-AG distance, as well as the sequence, is important for step II. In support of this conclusion, pyK, which has a BPS-to-AG distance of only 38 nt, undergoes step II more efficiently than all of the pyrimidine insertion mutants with BPS-to-AG distances of 49 nt (Figure 7C and D). Moreover, pre-mRNAs with BPS-to-AG distances of only 23 nt and pyrimidine tracts interrupted with purines or cytosines undergo catalytic step II as efficiently as WT (data not shown). Together, these data indicate that both the distance and the sequence between the BPS and AG play critical roles in catalytic step II of the splicing reaction. The further the AG is from the BPS, the more critical the BPS-to-AG sequence.

Discussion

We have identified the first mammalian spliceosomal proteins that interact with the pyrimidine tract at the 3' splice site prior to catalytic step II of the splicing reaction. We provide evidence that three of these proteins, of 110, 116 and 220 kDa, are components of U5 snRNP, indicating a central role for this snRNP in 3' splice-site recognition for catalytic step II in mammals. In most pre-mRNAs, the first AG downstream of the BPS is chosen as the 3' splice site. Our data establish the importance of two key parameters for selection of this AG in mammals, the presence of pyrimidines upstream of the AG and the distance between the BPS and the AG. These observations, together with our biochemical data on the RNA-protein interactions spanning from the BPS to the 3' splice site, suggest that AG selection occurs via a mechanism involving contacts (direct or indirect) between U5 snRNP bound at the pyrimidine tract and U2 snRNP bound at the BPS. A number of recent studies in yeast (Byrs and Schwer, 1996; Lauber *et al.*, 1996; Xu *et al.*, 1996; Luukkonen and Seraphin, 1997) suggest that a similar network of interactions may also be involved in 3' splice-site selection in yeast (see below).

RNA-protein interactions in the region between the BPS and AG

Using ³²P site-specific labeling and UV crosslinking, we found that 110, 116 and 220 kDa proteins crosslink to the 3' portion of the intron in the spliceosomal complex C. We note that although the C complex is usually defined as a single complex, it is actually highly dynamic, undergoing multiple changes during the two catalytic steps of the splicing reaction. Our data indicate that the 110, 116 and 220 kDa proteins crosslink to wild-type AdML pre-mRNA, as well as to an AdML pre-mRNA containing a GG substitution of the AG dinucleotide, a mutation that abolishes catalytic step II. Because high levels of the products of step I accumulate on this mutant, we used it for most of our analyses.

To determine how the 110, 116 and 220 kDa proteins are arranged on the 3' splice site, we labeled AdML pre-mRNA at several sites and analyzed the patterns of crosslinked proteins (see model, Figure 8A). Proceeding from 5' to 3', the U2 snRNP protein SAP 155 crosslinks

Fig. 8. A model for selection of the 3' splice site. (A) RNA-protein interactions in the C complex (only the factors relevant to this study are indicated). The branchpoint sequence (BPS), pyrimidine tract (py tract) and AG at the 3' splice site are indicated. The BPS-to-AG distance for wild-type AdML pre-mRNA is shown. Boxes and lines represent the exons and intron, respectively. The double-headed arrow indicates possible interactions between U2 and U5 snRNPs. (B) Temporal order of interactions during spliceosome assembled. Designations are as in (A), and only the interactions relevant to this study are shown (see text). As indicated, the same proteins are seen on the 3' splice site containing an AG or a GG, except for AG⁷⁵ which only crosslinks to pre-mRNA when an AG is present. Note that py¹⁴⁰ is only shown in the A complex, but is detected at variable levels in the B and C complexes (see text).

on both sides of the branch site and is the only protein detected at a site 5 nt downstream from the branch site (O.Gozani and R.Reed, unpublished observations). The relative levels of the 116 kDa protein are the highest 6 nt further downstream, while the 110 kDa protein and 220 kDa proteins are detected the strongest at a site 7 nt beyond. We were unable to determine the relative locations of the 110 and 220 kDa proteins, either because they bind very closely to each other on the pre-mRNA or bind to the same site sequentially (see Results). Little crosslinking of any proteins is detected directly at the GG, but on

wild-type pre-mRNA, crosslinking of a 75 kDa protein (designated AG⁷⁵) is detected at the AG (this study; also Chiaira *et al.*, 1996). The highly conserved loop in U5 snRNA crosslinks to position -1 of exon 1 and to position +1 of exon 2 prior to catalytic step II (Sontheimer and Steitz, 1993). Finally, at a site 7 nt downstream of the AG (Chiaira *et al.*, 1996) or 6 nt downstream of the GG (this study), crosslinking of the 220 kDa protein is detected. Our data also show that a region of pre-mRNA between the BPS and the AG is specifically protected from nuclease digestion in the C complex. These data, together with the crosslinking analyses, suggest that RNA-protein interactions span the distance from the BPS to just downstream of the 3' splice site.

Parameters for efficient AG selection

Previous studies of the mechanism for AG selection have led to a scanning model in which a factor binds at the BPS and scans to the nearest AG, without a strict distance or sequence requirement (Smith *et al.*, 1989, 1993). Thus, we were surprised to find a significant decrease in catalytic step II efficiency when an AG was moved as little as 6 nt downstream from a location where efficient AG use occurs. This observation is difficult to reconcile with a simple scanning mechanism and thus prompted us to look systematically at the role of BPS-to-AG distance and sequence in AG selection. We constructed pre-mRNAs in which a stretch of ~25 random nucleotides was inserted between the pyrimidine tract and the AG, increasing the BPS-to-AG distance from 23 nt in wild-type to ~50–60 nt in the mutants. Strikingly, these insertions abolish catalytic step II; the same result was observed with three completely different random sequences, making it unlikely that the effect is due to an inhibitory element or to formation of an inhibitory structure. Significantly, catalytic step II was partially rescued by replacement of the random nucleotides with random pyrimidines. In total, 16 different random pyrimidine insertions were tested (this study and data not shown). We conclude that the presence of pyrimidines adjacent to the AG plays an important role in AG use.

Although our data show that the sequence between the BPS and AG plays a critical role in catalytic step II, none of the pyrimidine insertion mutants undergoes catalytic step II as efficiently as wild-type pre-mRNA. All of these mutants have a BPS-to-AG distance of ~50 nt. Significantly, when this distance is decreased to 38 nt, step II efficiency is increased, and the highest level of step II is observed with the wild-type BPS-to-AG distance of 23 nt. Together, these data indicate that there is a maximal BPS-to-AG distance for efficient step II. Consistent with previous work (Smith *et al.*, 1989, 1993), we also find that there is a minimal BPS-to-AG distance. Specifically, an AG located 11 nt from the BPS is skipped in favor of a downstream AG, whereas an AG located 15 nt from the BPS is used.

Model for AG selection

As indicated above, the BPS-to-AG distance plays a key role in catalytic step II. One explanation for this distance constraint is that the RNA-protein interactions that span the distance from the BPS to the AG in the C complex delimit the boundaries for efficient AG selection (see Figure 8A). In addition, the presence of pyrimidines is

required for efficient catalytic step II. As the 110, 116 and 220 kDa proteins crosslink to the pyrimidine tract in the C complex, it is likely that one or more of these interactions also explain this constraint. Several observations suggest that the 110, 116 and 220 kDa crosslinked proteins are components of U5 snRNP. The data are the most compelling for the 220 kDa protein, which co-migrates with U5²²⁰ on 2D gels and is immunoprecipitated by antibodies to this protein. Although the 110 and 116 kDa proteins co-migrate on 2D gels with U5¹¹⁰ and U5¹¹⁶, respectively, we do not have data that definitively prove these correspondences (see Results). Previous biochemical depletion/add-back studies in both mammals and yeast have shown that U5 snRNP plays a role in catalytic step II (Winkelman *et al.*, 1989; Lamm *et al.*, 1991; Seraphin *et al.*, 1991), consistent with our conclusions. Moreover, PRP8, the yeast homolog of U5²²⁰ crosslinks to the 3' splice site prior to catalytic step II (Umen and Guthrie, 1995b). Although it is not yet known if the crosslinking is to the uridine tract, a mutant allele of PRP8 is impaired for recognition of the uridine tract and inhibits catalytic step II (Umen and Guthrie, 1995b, 1996).

On the basis of the correlation between the BPS-to-AG sequence/distance requirements and the RNA-protein interactions spanning from the BPS to AG, we propose that selection of the first AG downstream from the BPS involves a mechanism in which the U5 snRNP proteins are positioned by interactions with the pyrimidine tract and with U2 snRNP at the BPS (see Figure 8A). The AG closest to the 3' boundary of the U5 snRNP binding site is in turn selected as the 3' splice site. Selection of the AG itself may involve an interaction between a U5 snRNP protein and an AG recognition factor, such as AG⁷⁵ (Chiaira *et al.*, 1996). According to the model, efficient AG selection will occur if the AG is located in close vicinity of the U5²²⁰ interaction site (Figure 8A). If the BPS-to-AG distance is increased, less efficient AG selection is expected (and observed). If the AG is too close to the BPS, steric hindrance from the U5 snRNP proteins is expected to inhibit the second step, and our crosslinking data support this prediction. As the AG is not required to position U5 snRNP, it is likely that both the BPS and pyrimidine tract play a role in this process. Although U5 snRNP interacts with the pyrimidine tract, this sequence element alone does not contain sufficient information to position the snRNP in the correct place at the correct time. The U2 snRNP protein SAP 155, which crosslinks on both sides of the branch site (O.Gozani and R.Reed, unpublished observations) is the closest protein (detectable by crosslinking) to the U5 snRNP proteins. Thus, it is likely that a direct or indirect interaction between U5 snRNP and SAP 155 and/or other U2 snRNP proteins plays a role in positioning U5 snRNP on the 3' splice site.

The proposed model can accommodate inefficient use of AGs located as far downstream as 180 nt or AGs located next to non-pyrimidine-rich sequences. In the case of distant AGs, U5 snRNP tethered to the branch site (via U2 snRNP) interacts at a low frequency with sequences next to the distant AG, looping out the intervening RNA. For AGs located next to non-pyrimidine-rich sequences, the U5 snRNP interaction with U2 snRNP may be sufficiently strong to obviate the pyrimidine requirement. In support of this notion, the requirement for pyrimidines

next to the AG is greater the further the AG is from the branch site (this study and unpublished results). The proposed model also readily explains why two CAGs located next to each other are used with similar efficiency (Scadden and Smith, 1995). In this case, both AGs are essentially the same distance from the BPS and thus the positioning of U5 snRNP would not be expected to discriminate between them.

We note that the BPS-to-AG sequence/distance is highly unlikely to be the sole determinant of catalytic step II efficiency. It is possible that sequences/structures at the 5' splice site, BPS, in the intron or in exons 1 or 2 affect step II (see Umen and Guthrie, 1995a for review). Thus, distinctive BPS-to-AG sequence/distance requirements for different pre-mRNAs may result from contributions by these other sequence elements.

Yeast versus metazoans

Our conclusion that U5 snRNP is a key factor for 3' splice-site recognition during catalytic step II of the splicing reaction correlates very well with previous genetic and crosslinking studies in yeast (see Introduction). Several recent studies in yeast are also consistent with the proposal that an interaction between U2 and U5 snRNPs plays a key role in catalytic step II. Luukkonen and Seraphin (1997) have found that in yeast, as in mammals, the BPS-to-AG distance is critical for step II. Significantly, their study revealed that the optimal distance is 18–22 nt, which correlates well with our finding that second-step efficiency decreases when the BPS-to-AG distance is increased from 23 nt to ~35–50 nt. Further support for the idea that an interaction between U2 and U5 snRNPs is involved in AG selection for step II comes from the identification of genetic interactions between U2 and U5 snRNP-step II components. Notably, Xu and co-workers (1996) identified Slt22 in a synthetic lethal screen with a mutant U2 snRNA. Lauber and co-workers (1996) also identified this protein (designated SNU246) and showed that it is the yeast homolog of a mammalian U5 snRNP protein (U5²⁰⁰) likely required for the second step of splicing. (Slt22/SNU246 also corresponds to a mutant BRR2, which was isolated in a cold-sensitive screen; Noble and Guthrie, 1996.) Furthermore, several step II proteins in yeast (SLU7, PRP8 and PRP17) were recently found to be synthetically lethal with U2 snRNA (Xu *et al.*, 1996). Thus our data, together with these findings in yeast, provide compelling evidence that interactions between U2 snRNP bound at the BPS and U5 snRNP bound at the 3' splice site play key roles in 3' splice-site selection for catalytic step II.

Temporal order of RNA–protein interactions

Our data, together with previous studies, reveal a distinct temporal order of RNA–protein interactions on the 3' portion of the intron (see model, Figure 8B). U2AF⁶⁵ first binds tightly to the pyrimidine tract (Zamore and Green, 1989) in the E complex (Bennett *et al.*, 1992) and then becomes less tightly bound during the E to A complex transition; U2AF⁶⁵ also becomes phosphorylated in the A complex, which may play a role in the altered binding strength (Champion-Arnaud *et al.*, 1995). Py¹⁴⁰ is a novel protein that first crosslinks to the pre-mRNA in the A complex. As this protein is detected at variable levels in

the different complexes, we do not know whether or not it remains bound throughout spliceosome assembly. Both U5 snRNA and U5²²⁰ crosslink to exon sequences immediately upstream of the 5' splice site in the B complex (Newman and Norman, 1991; Wassarman and Steitz, 1992; Wyatt *et al.*, 1992; Cortes *et al.*, 1993; Sontheimer and Steitz, 1993; Chiara *et al.*, 1996; O'Keefe *et al.*, 1996). In addition, the 116 kDa protein first crosslinks to the pyrimidine tract in the B complex. Because our B complex preparation is contaminated with the A complex, we do not know whether U2AF⁶⁵ remains bound when the 116 kDa protein crosslinks. A major decrease in U2AF⁶⁵ crosslinking is detected in the C complex, concomitant with the crosslinking of the 110 kDa protein and U5²²⁰. Our observation that U5²²⁰ crosslinks to exon sequences downstream of the 3' splice site supports the previous proposal that U5²²⁰ together with U5 snRNA may function to hold the exons together for ligation (Wyatt *et al.*, 1992; Sontheimer and Steitz, 1993; Teigelkamp *et al.*, 1995). Finally, AG⁷⁵, which crosslinks very near or at the AG, may interact with the pre-mRNA after U5 snRNP binds, as AG⁷⁵, unlike U5 snRNP, requires a functional AG for crosslinking (Chiara *et al.*, 1996).

Materials and methods

Plasmids

WT AdML is encoded by pAdML which was described (Michaud and Reed, 1993). GG-AG, GG---AG and GG-GG are identical to AdML except for the sequences indicated in the figures. pyA-pyK and ranB-ranC differ from AdML by sequences indicated in the figures and by insertion of the four nucleotides GCUC immediately downstream of the 3' splice site. WT' is the same as WT AdML except for the GCUC insertion immediately downstream of the 3' splice site. 15AG and 11AG pre-mRNAs are isogenic with AdML except between the BPS and the second downstream AG, as indicated in the figure. The plasmids GG-AG, 15AG, 11AG, pyA-pyK and ranB-ranC were constructed by cloning the appropriate oligonucleotides into the *Hind*III and *Pst*I sites of pAdML. GG-GG was constructed by cloning appropriate oligonucleotides into the *Hind*III and *Sal*I sites of pAdML. The plasmid encoding GG---AG pre-mRNA was constructed by ligating oligonucleotides into the *Pst*I site of plasmid GG-AG. RanA pre-mRNA is identical to pAdMLΔAG (described in Gozani *et al.*, 1994) except for the inserted sequence indicated in the figure, and it was constructed by ligating oligonucleotides into the *Nco*I site of pAdMLΔAG. All DNAs were linearized with *Bam*HI and transcribed with T7 RNA polymerase.

Site-specific labeling and UV crosslinking

³²P-site-specifically labeled pre-mRNAs were synthesized as described (Moore and Sharp, 1992; Chiara *et al.*, 1996). Guanosine residues were chosen for site-specific labeling because transcription initiates most efficiently with this nucleotide. Where necessary, a G residue was introduced into the site used for labeling. For the I⁻¹³ site, it was necessary to change the UU at I⁻¹³ to GC in order to transcribe the RNA. Spliceosomal complexes A/B and C were assembled on WT AdML pre-mRNA by incubating splicing reactions at 30°C for 5 and 30 min, respectively (note that it is not possible to obtain significant amounts of the A complex without some contaminating B complex). Assembly of the ATP-independent E complex on WT AdML was performed as previously described (Michaud and Reed, 1993). For assembly of A/B, B and C complexes on GG-GG pre-mRNA, reaction mixtures were incubated at 30°C for 7.5, 15 and 70 min, respectively. Spliceosomal complexes were isolated by gel filtration and UV crosslinked (Gozani *et al.*, 1996). Affinity purification of spliceosomal complex C and UV crosslinking were performed as described (Gozani *et al.*, 1996). RNase A (8 µg) or RNase T1 (4–10 µg) was added to 300 µl aliquots of the gel filtration fractions which were then incubated for 30 min at 37°C. Proteins were analyzed on SDS or on 2D gels. RNA from gel filtration fractions was fractionated on 15% denaturing polyacrylamide gels.

Immunoprecipitation of crosslinked proteins

15 µl of α-p220 was coupled to 40 µl of protein A–Trisacryl beads. 500 µl of mAb386 or U1 70K was coupled to 40 µl protein A–Trisacryl beads by using 40 µg of goat anti-mouse secondary antibody (Behrens and Luhrmann, 1991). The coupled beads were mixed with a 1 ml aliquot of gel filtration fraction containing the C complex assembled on GG-GG pre-mRNA labeled at 1⁻⁶; the mixture was rotated overnight at 4°C. Prior to mixing with the Trisacryl, the gel filtration fraction was irradiated with UV light (as above), treated with RNase A (12 µg/ml), then proteins were either used directly for immunoprecipitation or denatured first. Denaturing was carried out by adding SDS to a final concentration of 1%, incubating at 70°C for 5 min, then adding Triton X-100 to a final concentration of 5%.

Acknowledgements

We are indebted to G.Moreau and M.Moore for providing p220 antisera and to R.Luhrmann for mAb386. We are grateful to K.Lynch for comments on the manuscript and members of the laboratory for useful discussions. This work was supported by a Tobacco Research Council grant and an NIH grant to R.R.

References

- Adams,M.D., Rudner,D.Z. and Rio,D.C. (1996) Biochemistry and regulation of pre-mRNA splicing. *Curr. Opin. Cell Biol.*, **8**, 331–339.
- Bach,M., Winkelmann,G. and Luhrmann,R. (1989) 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl Acad. Sci. USA*, **86**, 6038–6042.
- Behrens,S.E. and Luhrmann,R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev.*, **5**, 1439–1452.
- Bennett,M., Michaud,S., Kingston,J. and Reed,R. (1992) Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev.*, **6**, 1986–2000.
- Billings,P.B., Allen,R.W., Jensen,F.C. and Hoch,S.O. (1982) Anti-RNP monoclonal antibodies derived from a mouse strain with lupus-like autoimmunity. *J. Immunol.*, **128**, 1176–1180.
- Byers,A. and Schwer,B. (1996) Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and 3' splice site. *RNA*, **2**, 707–717.
- Champion-Arnaud,P., Gozani,O., Palandjian,L. and Reed,R. (1995) Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol. Cell. Biol.*, **15**, 5750–5756.
- Chiara,M.D., Gozani,O., Bennett,M., Champion-Arnaud,P., Palandjian,L. and 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.
- Cortes,J.J., Sontheimer,E.J., Seiwert,S.D. and Steitz,J.A. (1993) Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5' splice sites *in vivo*. *EMBO J.*, **12**, 5181–5189.
- Frank,D. and Guthrie,C. (1992) An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.*, **6**, 2112–2124.
- Frank,D., Patterson,B. and Guthrie,C. (1992) Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. *Mol. Cell. Biol.*, **12**, 5197–5205.
- Gozani,O., Patton,J.G. and Reed,R. (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.
- Gozani,O., Feld,R. and Reed,R. (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.
- Horowitz,D.S. and Krainer,A.R. (1997) A human protein required for the second step is functionally related to a yeast splicing factor. *Genes Dev.*, **11**, 139–151.
- Kandels-Lewis,S. and Seraphin,B. (1993) Role of U6 snRNA in 5' splice site selection. *Science*, **262**, 2035–2039.
- Kramer,A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, **65**, 367–409.
- Lamm,G.M., Blencowe,B.J., Sproat,B.S., Iribarren,A.M., Ryder,U. and Lamond,A.I. (1991) Antisense probes containing 2-aminoadenosine allow efficient depletion of U5 snRNP from HeLa splicing extracts. *Nucleic Acids Res.*, **19**, 3193–3198.
- Lauber,J., Fabrizio,P., Teigelkamp,S., Lane,W.S., Hartmann,E. and Luhrmann,R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J.*, **15**, 4001–4015.
- Lesser,C.F. and Guthrie,C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982–1988.
- Lossky,M., Anderson,G.J., Jackson,S.P. and Beggs,J. (1987) Identification of a yeast snRNP protein and detection of snRNP–snRNP interactions. *Cell*, **51**, 1019–1026.
- Luukkainen,B.G.M. and Seraphin,B. (1997) The roles of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J.*, **16**, 779–792.
- Madhani,H.D. and Guthrie,C. (1994) Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1–26.
- Michaud,S. and Reed,R. (1993) A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.*, **7**, 1008–1020.
- Moore,M.J. and Sharp,P.A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, **256**, 992–997.
- Neugebauer,K.M., Stolk,J.A. and Roth,M.B. (1995) A conserved epitope on a subset of SR proteins defines a larger family of pre-mRNA splicing factors. *J. Cell Biol.*, **129**, 899–908.
- Newman,A.J. (1994) Pre-mRNA splicing. *Curr. Opin. Genet. Dev.*, **4**, 298–304.
- Newman,A. and Norman,C. (1991) Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell*, **65**, 115–123.
- Newman,A.J. and Norman,C. (1992) U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell*, **68**, 743–754.
- Newman,A.J., Teigelkamp,S. and Beggs,J.D. (1995) snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA*, **1**, 968–980.
- Nilsen,T.W. (1994) RNA–RNA interactions in the spliceosome: unraveling the ties that bind. *Cell*, **78**, 1–4.
- Noble,S.M. and Guthrie,C. (1996) Identification of novel genes required for yeast pre-mRNA splicing by means of cold sensitive mutations. *Genetics*, **143**, 67–80.
- O'Keefe,R.T., Norman,C. and Newman,A.J. (1996) The invariant U5 snRNA loop I sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell*, **86**, 679–689.
- Parker,R., Siliciano,P.G. and Guthrie,C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, **49**, 229–239.
- Patterson,B. and Guthrie,C. (1991) A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell*, **64**, 181–187.
- Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.*, **7**, 393–406.
- Reed,R. (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113–2123.
- Reed,R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.*, **6**, 215–220.
- Sawa,H. and Abelson,J. (1992) Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. *Proc. Natl Acad. Sci. USA*, **89**, 11269–11273.
- Sawa,H. and Shimura,Y. (1992) Association of U6 snRNA with the 5' splice site region of pre-mRNA in the spliceosome. *Genes Dev.*, **6**, 244–254.
- Scadden,A.D.J. and Smith,C.W.J. (1995) Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.*, **14**, 3236–3246.
- Seraphin,B., Abovich,N. and Rosbash,M. (1991) Genetic depletion indicates a late role for U5 snRNP during *in vitro* spliceosome assembly. *Nucleic Acids Res.*, **19**, 3857–3860.
- Smith,C.W., Porro,E.B., Patton,J.G. and Nadal-Ginard,B. (1989) Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature*, **342**, 243–247.
- Smith,C.W., Chu,T.T. and Nadal-Ginard,B. (1993) Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.*, **13**, 4939–4952.
- Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.

- Spritz,R.A., Strunk,K., Surowy,C.S., Hoch,S.O., Barton,D.E. and Franke,U. (1987) The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res.*, **15**, 10373–10390.
- Staknis,D. and Reed,R. (1994) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol. Cell. Biol.*, **14**, 2994–3005.
- Staknis,D. and Reed,R. (1995) Members of a family of proteins (the RD family) detected by a U1 70K monoclonal antibody are present in spliceosomal complexes. *Nucleic Acids Res.*, **23**, 4081–4086.
- Teigelkamp,S., Newman,A.J. and 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.
- Umen,J.G. and Guthrie,C. (1995a) The second catalytic step of pre-mRNA splicing. *RNA*, **1**, 869–885.
- Umen,J.G. and Guthrie,C. (1995b) A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.*, **9**, 855–868.
- Umen,J.G. and Guthrie,C. (1995c) Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA*, **1**, 584–597.
- Umen,J.G. and Guthrie,C. (1996) Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics*, **143**, 723–739.
- Wassarman,D.A. and Steitz,J.A. (1992) Interactions of small nuclear RNAs with precursor messenger RNA during *in vitro* splicing. *Science*, **257**, 1918–1925.
- Whittaker,E., Lossky,M. and Beggs,J.D. (1990) Affinity purification of spliceosomes reveals that the precursor RNA processing protein PRP8, a protein in the U5 small nuclear ribonucleoprotein particle, is a component of yeast spliceosome. *Proc. Natl Acad. Sci. USA*, **87**, 2216–2219.
- Winkelman,G., Bach,M. and Luhrmann,R. (1989) Evidence from complementation assays *in vitro* that U5 snRNP is required for both steps of mRNA splicing. *EMBO J.*, **8**, 3105–3112.
- Wu,J. and Manley,J.L. (1989) Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.*, **3**, 1553–1561.
- Wyatt,J.R., Sontheimer,E.J. and 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.
- Xu,D., Nouraini,S., Field,D., Tang,S. and Friesen,J.D. (1996) An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. *Nature*, **381**, 709–713.
- Zamore,P.D. and Green,M.R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA*, **86**, 9243–9247.
- Zhuang,Y. and Weiner,A.M. (1989) A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev.*, **3**, 1545–1552.

Received on March 20, 1997; revised on May 6, 1997