Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A

Or Gozani, Rebecca Feld, and Robin Reed
Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115 USA

A critical step in the pre-mRNA splicing reaction is the stable binding of U2 snRNP to the branchpoint sequence (BPS) to form the A complex. The multimeric U2 snRNP protein complexes SF3a and SF3b are required for A complex assembly, but their specific roles in this process are not known. *Saccharomyces cerevisiae* homologs of all of the SF3a, but none of the SF3b, subunits have been identified. Here we report the isolation of a cDNA encoding the mammalian SF3b subunit SAP 145 and the identification of its probable yeast homolog (29% identity). This first indication that the homology between yeast and metazoan A complex proteins can be extended to SF3b adds strong new evidence that the mechanism of A complex assembly is highly conserved. To investigate this mechanism in the mammalian system we analyzed proteins that cross-link to 32P-site-specifically labeled pre-mRNA in the A complex. This analysis revealed that SAP 145, together with four other SF3a/SF3b subunits, UV cross-links to pre-mRNA in a 20-nucleotide region upstream of the BPS. Mutation of this region, which we have designated the anchoring site, has no apparent effect on U2 snRNP binding. In contrast, when a 2'-O methyl oligonucleotide complementary to the anchoring site is added to the spliceosome assembly reaction, A complex assembly and cross-linking of the SF3a/SF3b subunits are blocked. These data indicate that sequence-independent binding of the highly conserved SF3a/SF3b subunits upstream of the branch site is essential for anchoring U2 snRNP to pre-mRNA.

[Key Words: pre-mRNA splicing, *S. cerevisiae* homolog, branchpoint sequence, snRNP protein complex, spliceosomal complex A; UV cross-links]

Received August 25, 1995; revised version accepted November 17, 1995.

During pre-mRNA splicing, a series of highly dynamic spliceosomal complexes assemble in the order E → A → B → C. Assembly of these complexes on pre-mRNA involves the functions of >50 distinct protein components as well as five small nuclear RNAs ([U1, U2, U4, U5, and U6 snRNAs]) [for review, see Moore et al. 1993, Newman 1994]. One of the key steps in the spliceosome assembly pathway is the formation of an essential base-pairing interaction between U2 snRNA and the branchpoint sequence [BPS]. This interaction is established in the A complex and is thought to position the branch-site adenosine for nucleophilic attack on the 5' splice site during catalytic step I of the splicing reaction [Query et al. 1994]. The BPS is highly conserved in yeast (consensus, UACUAAC), but only weakly conserved in metazoans (consensus, YNRAy). Given the short length of the BPS, and its degeneracy in metazoans, it is likely that spliceosomal proteins function together with the U2 snRNA–BPS duplex to tether the snRNP to pre-mRNA during spliceosome assembly.

In metazoans, affinity-purified A complex consists primarily of U2 snRNP-specific proteins; these include A' and B'', as well as spliceosome-associated proteins [SAPs] 49, 61, 62, 114, 130, 145, and 155 [for review, see Hodges and Beqgs 1994]. SAPs 61, 62, and 114 correspond to subunits of the multimeric splicing factor SF3a [SF3a°, SF3a°, and SF3a°, respectively], and SAPs 49, 130, 145, and 155 are thought to be subunits of another splicing factor, SF3b [SF3b°, SF3b°, and SF3b°, respectively] [Brosi et al. 1993a,b]. In addition, all of the SF3a and SF3b subunits, except SAP 130, can be UV cross-linked to pre-mRNA in affinity-purified spliceosomal complexes A, B, and C [Gozani et al. 1994; Staknis and Reed 1994]. The site of SAP 49 cross-linking in the A complex was localized to within a 29-nucleotide region in the vicinity of the BPS [Champion-Arnaud et al. 1994], and the cross-linking sites of the other SF3a/SF3b subunits are not known.
The protein composition of the yeast A complex has not been determined. However, this complex has been shown to contain all three subunits of SF3a; the essential yeast pre-mRNA processing (PRP) factors PRP9, PRP11, and PRP21 correspond to SAPs 61, 62, and 114 (for reviews, see Hodges and Beggs 1994; Newman 1994). cDNAs encoding SAPs 61 (Chiara et al. 1994; Kramer et al. 1994), 62 [Bennett and Reed 1993], and SF3a120 [Kramer et al. 1995] have been isolated and the deduced amino acid sequences found to be ~25% identical to the corresponding yeast factors. Although it is not yet known whether the yeast SF3a homologs UV cross-link to pre-mRNA, the protein–protein interactions between the SF3a subunits appear to be conserved. In mammals, SAPs 61 and 62 both interact with SAP 114, while in yeast PRPs 9 and 11 both interact with PRP21 [Bennett and Reed 1993; Legrain and Chapon 1993; Legrain et al. 1993; Chiara et al. 1994; Kramer et al. 1994, 1995]. Despite the apparent conservation between yeast and mammal in assembly of the A complex, no yeast homologs of any of the SF3b subunits have been reported.

Here we report the isolation and characterization of a cDNA encoding the SF3b subunit SAP 145. Comparison of the predicted SAP 145 amino acid sequence to sequences in the data base revealed a Saccharomyces cerevisiae gene encoding a protein that is 29% identical to SAP 145. This yeast gene is the same as CUS1, a gene isolated as a suppressor of a cold-sensitive mutation in U2 snRNA [as shown in Wells et al. (this issue)]. Thus, the CUS1 protein is the probable yeast homolog of SAP 145. SAP 145 interacts directly with SAP 49 [Champion-Arnaud and Reed 1994], and another S. cerevisiae gene in the data base encodes a protein that is 36% identical to SAP 49 [for alignment, see Wells et al. (this issue)]. The conservation from yeast to humans of all three SF3a subunits, and at least two of the SF3b subunits (SAPs 49 and 145), implicates these proteins in critical steps of the splicing reaction. Consistent with this prediction, our data show that the mammalian SF3a/SF3b subunits UV cross-link to a specific region of pre-mRNA upstream of the branch site in functional A complex. Moreover, our data provide evidence that one or more of these RNA–protein interactions function to anchor U2 snRNP to pre-mRNA during spliceosome assembly.

Results

Isolation of a cDNA encoding SAP 145

To isolate a cDNA encoding SAP 145, the A complex was purified in large scale, fractionated on two-dimensional gels and peptide sequence information was obtained from tryptic digests of SAP 145 [see Materials and methods; Bennett and Reed 1993]. Degenerate oligonucleotides were designed to a tryptic peptide and used as PCR primers to isolate a partial cDNA encoding SAP 145 from HeLa cell poly(A)+ mRNA. This cDNA was then used as a probe to screen a cDNA library and isolate a full-length SAP 145 cDNA. The in vitro translation product of the SAP 145 cDNA precisely cofractionates with SAP 145 present in the A complex (Fig. 1B). In addition, the predicted initiator methionine is preceded by a translation initiation consensus sequence [Kozak 1986; data not shown]. Finally, both of the peptide sequences obtained from SAP 145 are present in the predicted SAP 145 amino acid sequence [see Materials and methods]. We conclude that the SAP 145 cDNA is full length. The SAP 145 cDNA predicts an 872 amino acid protein with a calculated molecular mass of 97.6 kD (Fig. 1A). Thus, SAP 145 migrates anomalously on SDS gels with an estimated size of 145 kD on a 9% gel and 120 kD on a 6% gel [Staknis and Reed 1994]. The protein is highly hydrophilic, the amino terminus is proline-rich, and the carboxyl terminus has a glutamate-rich stretch [Fig. 1A]. In contrast to SAPs 49, 61, and 62, no known RNA-binding motifs are present in SAP 145.

Comparison of the SAP 145 amino acid sequence to the GenBank data base revealed a cosmid sequence that contains a gene encoding a potential S. cerevisiae homolog of SAP 145 [accession no. Z48,756]. This gene, with no reported function, predicts a 436-amino-acid protein. An alignment of the region of highest similarity between SAP 145 and the yeast protein, designated ySAP 145, is shown in Figure 1C. Within this region, ySAP 145 and SAP 145 are 29% identical, and there is a 150-amino-acid stretch [Fig. 1C; SAP 145 amino acids 477–631] where the identity is 47%. The identity between SAP 145 and ySAP 145 is equal to, or greater than, that found between the yeast and mammalian SF3a subunits [Bennett and Reed 1993; Chiara et al. 1994; Kramer et al. 1994, 1995].

SF3a/SF3b components UV cross-link upstream of the branch site

The identification of ySAP 145 provided the first indication that SF3b is conserved between yeast and mammals and implicated a central role for this factor in A complex assembly. SAPs 145 and 49 interact directly, and SAP 49 UV cross-links to pre-mRNA in the vicinity of the BPS (Champion-Arnaud and Reed 1994). To understand the function of SAPs 145 and 49, we first sought to determine whether SAP 145 also cross-links near this critical sequence element (the BPS) and to determine more precisely the cross-linking site of SAP 49. Previous cross-linking studies have used photoactivatable nucleotide analogs, such as 4-thiouridine [Wyatt et al. 1992, Gaur et al. 1995] or benzophenone-derivatized adenosine [MacMillan et al. 1994]. The advantage of these derivatized nucleotides is their high cross-linking efficiency. The disadvantages are that only a particular nucleotide can be used [i.e., U in the case of 4-thiouridine] and that the derivatized nucleotide can decrease the efficiency of spliceosome assembly. In addition, in the case of benzophenone-derivatized adenosine, all proteins within a 15 Å radius of the nucleotide are cross-linked [MacMillan et al. 1994]. To circumvent these difficulties, we synthesized 32P-site-specifically labeled pre-mRNA and carried out UV cross-linking of gel filtration-isolated complexes. Because the complexes are partially purified,
Role of highly conserved U2 snRNP proteins

Figure 1. Isolation of a cDNA encoding SAP 145 and alignment with probable yeast homolog. (A) Predicted amino acid sequence of SAP 145. Amino-terminal proline-rich and carboxy-terminal glutamate-rich stretches are underlined. (B) Alignment of SAP 145 and probable yeast homolog, ySAP 145. Amino acids 377 through the carboxy-terminal residue (amino acid 8721 of SAP 145 are aligned with amino acids 83 through the carboxy-terminal residue (amino acid 436j of ySAP 145. Identical residues are designated by two dots and similarities by a single dot. Amino acid groupings are as follows: (Q, E, N, D), (K, R), (L, I, V), (F, Y), (A, G), (S, T).

To determine whether SAP 145 UV cross-links to pre-mRNA near SAP 49, we site-specifically labeled the G residue 15 nucleotides upstream of the branch site (−15; Fig. 2). Four cross-linked bands were detected at the −15 site, one of which comigrates with in vitro translated (IVT) SAP 49 (Fig. 3A, lanes 1, 3). In addition, the highest molecular mass band comigrates with IVT–SAP 145 (Fig. 3A, lanes 2, 3). SAP 145 migrates as a 120-kD protein on a large amount of sample can be loaded on SDS gels, thus overcoming the inefficiency of UV cross-linking to unmodified RNA. Using our strategy, spliceosome assembly is not affected, and only proteins in direct contact

a large amount of sample can be loaded on SDS gels, thus overcoming the inefficiency of UV cross-linking to unmodified RNA. Using our strategy, spliceosome assembly is not affected, and only proteins in direct contact with the pre-mRNA are detected.

To determine whether SAP 145 UV cross-links to pre-mRNA near SAP 49, we site-specifically labeled the G residue 15 nucleotides upstream of the branch site (−15; Fig. 2). Four cross-linked bands were detected at the −15 site, one of which comigrates with in vitro translated (IVT) SAP 49 (Fig. 3A, lanes 1, 3). In addition, the highest molecular mass band comigrates with IVT–SAP 145 (Fig. 3A, lanes 2, 3). SAP 145 migrates as a 120-kD protein on a large amount of sample can be loaded on SDS gels, thus overcoming the inefficiency of UV cross-linking to unmodified RNA. Using our strategy, spliceosome assembly is not affected, and only proteins in direct contact with the pre-mRNA are detected.

To determine whether SAP 145 UV cross-links to pre-mRNA near SAP 49, we site-specifically labeled the G residue 15 nucleotides upstream of the branch site (−15; Fig. 2). Four cross-linked bands were detected at the −15 site, one of which comigrates with in vitro translated (IVT) SAP 49 (Fig. 3A, lanes 1, 3). In addition, the highest molecular mass band comigrates with IVT–SAP 145 (Fig. 3A, lanes 2, 3). SAP 145 migrates as a 120-kD protein on a large amount of sample can be loaded on SDS gels, thus overcoming the inefficiency of UV cross-linking to unmodified RNA. Using our strategy, spliceosome assembly is not affected, and only proteins in direct contact with the pre-mRNA are detected.

To determine whether SAP 145 UV cross-links to pre-mRNA near SAP 49, we site-specifically labeled the G residue 15 nucleotides upstream of the branch site (−15; Fig. 2). Four cross-linked bands were detected at the −15 site, one of which comigrates with in vitro translated (IVT) SAP 49 (Fig. 3A, lanes 1, 3). In addition, the highest molecular mass band comigrates with IVT–SAP 145 (Fig. 3A, lanes 2, 3). SAP 145 migrates as a 120-kD protein on a large amount of sample can be loaded on SDS gels, thus overcoming the inefficiency of UV cross-linking to unmodified RNA. Using our strategy, spliceosome assembly is not affected, and only proteins in direct contact with the pre-mRNA are detected.
slightly faster and more basic than the 49-kD cross-linked protein [the cross-linked RNA fragment on the protein causes a small shift of lower molecular mass cross-linked proteins (Staknis and Reed 1994)]. In addition, the two-dimensional gel data (Fig. 3B) indicate that the other two cross-linked bands (Fig. 3A, lane 3) correspond to SAPs 114, 62, and 61 [the correspondence between cross-linked and silver-stained SAP 114 is seen clearly on a longer exposure of the gel (data not shown)]. These data are consistent with studies showing that all of these U2 snRNP proteins (SAPs 49, 61, 62, 114, and 145) cross-link to pre-mRNA in affinity-purified spliceosomes (Staknis and Reed 1994).

On the basis of these data, we conclude that five of the SF3a and SF3b subunits cross-link to pre-mRNA upstream of the branch site. Cross-linking of these proteins is not detected when pre-mRNA is labeled at the 5' or 3' splice sites, in the exons, or directly at the branch site (O. Gozani, M. D. Chiara, and R. Reed, unpubl.). In addition, the same cross-linking pattern is observed when the premRNA is incubated in splicing extracts for different times during A complex assembly [e.g., 2.5 min (see Fig. 7C, below), 5, 10, or 12.5 min, data not shown]. Thus, it is unlikely that the detection of multiple proteins at a single site is attributable to heterogeneity in the A complex. Instead, our data (see below) indicate that multiple proteins are detected at a single nucleotide because a large number of proteins cross-link to a short region of the pre-mRNA and thus, protect partially the RNA from RNase digestion. Protection of this region from RNase digestion in the A complex has been reported previously [e.g., Kramer 1988]. Given this limitation, our strategy for establishing the cross-linking sites of each protein was to compare the relative levels of cross-linked proteins at different sites along the pre-mRNA after treating complexes with either RNase T1 or RNase A. In the data presented below, equal counts per minute of the A complexes assembled on the site-specifically labeled pre-mRNAs were loaded on the SDS gels. On the two-dimensional gels (in Figs. 3B, 4B, and 5A), only the relative
Role of highly conserved U2 snRNP proteins

Figure 4. Comparison of proteins that cross-link at −25, −15, and −12. (A) Cross-linked proteins detected when pre-mRNA is labeled at −25 (lanes 1, 4, 7), −15 (lanes 2, 5, 8), or −12 (lanes 3, 6, 9). Cross-linked A complexes were treated with RNase A or T1 as indicated and equal counts per minute were loaded on the gel. Lanes 7–9 are a darker exposure of lanes 4–6. (B) −12 or −25 cross-linked samples (treated with RNase A) were mixed with affinity-purified A complex and then fractionated on a two-dimensional gel. The identities of the cross-linked bands were determined by superimposing the cross-linking and silver-stained patterns (data not shown).

levels of cross-linked proteins can be compared because exposures were varied to best show the different proteins.

After RNase A digestion, SAP 49 is detected most strongly at position −25, whereas SAP 61 is the strongest at −15 [Fig. 4A, lanes 4–6]. In addition, SAPs 49, 61, and 114 are detected much more strongly at position −15 after RNase T1, than RNase A, digestion [Fig. 4A, cf. lanes 2 and 5; equal cpm's were loaded]. These results indicate that these three proteins cross-link on the large RNase T1 fragment located between −25 and −15 (underlined, see Fig. 2). It is likely that SAP 114 is between SAPs 49 and 61 on this RNase T1 fragment because SAP 114 is detected at similar levels at −25 and −15 [Fig. 4A, lanes 7–9], whereas SAP 49 is detected strongly only at −25 and SAP 61 only at −15 [Fig. 4A, lanes 4–6]. [The lower molecular mass proteins that cross-link at the −25 site [Fig. 4A, lane 7] are hnRNP proteins from a low level of contaminating H complex (data not shown).]

SAP 145 cross-linking is detected stronger at −15 and −12 versus −25 after RNase A digestion [Fig. 4A, lanes 7–9] and is not detected strongly on any RNase T1 fragments [Fig. 4A, lanes 1–3, and longer exposure of these lanes, data not shown]. These results indicate that SAP 145 is downstream of the G residue at −12 [see Fig. 2].

With label at −6 (see Fig. 2), we detect both SAPs 62 and 155 cross-linking [Fig. 5A; proteins were identified by superimposing cross-linking and silver-stained gel patterns (data not shown)]. SAP 62 is likely upstream of SAP 155 as SAP 62 is detected at both −12 and −15,
Figure 6. Mutation of the anchoring site does not affect A complex assembly or the apparent 5' to 3' linear cross-linking order of SF3a/SF3b subunits. (A) A complex assembly. The schematic shows the sequences of wild-type (wt) and mutant [a-e] anchoring sites. The bold nucleotides indicate the branchpoint sequence (BPS). Each of the pre-mRNAs (20 ng) was incubated under splicing conditions (25 µl reactions) for 3 min. and then fractionated on a non-denaturing gel. The bands corresponding to the A and H complexes are indicated. (B) SF3a/SF3b cross-linking. The schematic shows the sequences of wild-type [wt] and mutant [f-i] pre-mRNAs, the arrows indicate the G residues that were labeled on pre-mRNA f. The numbers are relative to the branch-site adenosine. The RNase T1 fragments [see text] are underlined, and the region of the pre-mRNA designated the anchoring site is shown. The bottom panel shows cross-linked proteins detected when wt pre-mRNA is labeled at −15 and when f pre-mRNA is labeled at −13 and −23. Cross-linked A complexes were treated with RNase A or T1 as indicated, and equal counts per minute were loaded on the gels. The gel on the left represents one experiment, and the two gels on the right are a second experiment (equal counts per minute loaded). We note that in this experiment high levels of SAP 145 were detected at the −15 site after RNase T1 digestion, most likely because of partial digestion with RNase T1 [because of a large volume for this particular sample]. This was not observed in other experiments (e.g., Figs. 4A and 7C).

whereas SAP 155 is detected barely at −12 and is not detected at all at −15 [see Figs. 3B and 4B]. SAP 62 is detected more strongly at the −6 site than at the −12 site [Fig. 5B] indicating that this protein is closer to the −6 site. SAP 145 is not detected at the −6 site, and therefore, is probably closer to the −12 site [Fig. 5].

On the basis of these data, we conclude that the SF3a/SF3b proteins UV cross-link to a 20-nucleotide region upstream of the branch site in the apparent 5' to 3' order SAPs 49, 114, 61, 145, 62, and 155 [see model, Fig. 8A below]. For simplicity, we will refer to this 20 nucleotide region as the anchoring site. In another study, we have found that SAP 155 also cross-links to a site 5 nucleotides downstream of the BPS; none of the other SF3a/SF3b subunits are detected downstream of the BPS, and SAP 155 cross-links on both sides of the BPS on two pre-mRNAs containing different sequences surrounding the BPS, arguing that this cross-linking pattern is functionally significant [O. Gozani and R. Reed, unpubl.].

Although our data are consistent with a model [Fig. 8A, below] in which the SF3a/SF3b subunits bind to pre-mRNA in a specific 5' to 3' linear order, we cannot rule out the possibility that some of the interactions detected by the cross-linking assay are nonfunctional. In addition, it is possible that some of the low efficiency cross-links are generated from proteins that are directly next to the pre-mRNA, but are not actually bound to it. Finally, UV-induced damage or UV-induced thermal motion of the proteins or RNA in the complex may result in detection of nonfunctional interactions. These possibilities notwithstanding, our data, as well as previous work [Ruskin and Green 1987, Kramer 1988; Ruskin et al. 1988], show that the pre-mRNA at the anchoring site is highly resistant to RNase digestion, consistent with the model that there are extensive RNA-protein interactions in this region. In addition, we observe a similar cross-linking pattern when UV irradiation is carried out for different lengths of time [data not shown], arguing that the UV irradiation itself does not damage or alter the complex under our conditions.

**SF3a/SF3b cross-linking pattern is independent of sequence**

Sequence comparisons of several different mammalian pre-mRNAs revealed that the 20-nucleotide anchoring site is pyrimidine rich [data not shown]. To determine whether there might be a sequence requirement, we synthesized pre-mRNAs in which the anchoring site was replaced with random sequences [Fig. 6A]. A comparison of the efficiency of A complex assembly [Fig. 6A] or splicing [data not shown] of the mutant and wild-type pre-mRNAs showed no apparent differences. Even pre-mRNA “d”, which is purine rich, does not differ significantly from wild type for spliceosome assembly and splicing [Fig. 6A; data not shown]. We conclude that there is no major sequence requirement at the anchoring site for spliceosome assembly or splicing.

To determine whether sequence changes in the anchoring site alter the cross-linking pattern of the SF3a/SF3b subunits, we examined the proteins that cross-link to pre-mRNA “i” in the A complex [Fig. 6B]. This pre-
mRNA, which assembles into A complex normally [data not shown], was chosen because it contains a large RNase T1 fragment at the anchoring site, thus facilitating analysis of the cross-linking pattern [Fig. 6B, top]. Pre-mRNA "f" was site-specifically labeled at positions −13 and −23, and wild-type pre-mRNA was labeled at position −15 for side-by-side comparison [Fig. 6B]. An example of data obtained after RNase A digestion of the cross-linked complexes is shown on the left, and a comparison of RNase A and T1 digests (equal cpms loaded) is shown in the middle and right [Fig. 6B, bottom].

Strikingly, these data reveal that the apparent linear cross-linking order of the SF3a/SF3b subunits is the same as that detected on wild-type pre-mRNA; the RNase A and T1 patterns at the −23 site on pre-mRNA "f" appear the same as the corresponding patterns at the −25 site on wild-type pre-mRNA [cf. Fig. 6B, −23 lanes with Fig. 4A, −25 lanes]. Likewise, the RNase A pattern at −13 on pre-mRNA "f" resembles the −12 pattern on wild type [cf. Fig. 6B, RNase A, −13 lanes with Fig. 4A, RNase A, −12 lanes]. In most preparations of the −13 sample (e.g., Fig. 6B, right panel −13), lower levels of cross-linked proteins are detected than are detected at the −12 site (on wild-type pre-mRNA). This is most likely because there are more RNase A digestion sites (i.e., pyrimidine residues) surrounding the −13 site (Fig. 6B top, cf. wild-type with "f" sequence). Finally, the pattern of cross-linking at the −13 site is consistent with the proposed cross-linking order [see model, Fig. 8A] as SAPs 145 and 61, which would be expected to be on either side of the −13 site, are detected strongly at this site after RNase T1 digestion [Fig. 6B, bottom, middle, −13]. Together, these data indicate that the apparent 5' to 3' linear cross-linking order of SF3a/SF3b is independent of the sequence at the anchoring site.

Role of SF3a/SF3b protein–pre-mRNA interactions in stable U2 snRNP binding

To determine whether the SF3a/SF3b protein–pre-mRNA interactions play a role in U2 snRNP binding, we synthesized a 2'O methyl oligonucleotide [Barabino et al. 1990] complementary to a 14-nucleotide sequence in the anchoring site [anchr, Fig. 7A]. A 14-nucleotide 2'O methyl oligonucleotide complementary to a region upstream of the anchoring site was used as a control [cntrl, Fig. 7A]. Significantly, we observed a concentration-dependent decrease in the efficiency of A complex assembly when different amounts of the anchr oligonucleotide were added to A complex assembly reactions [Fig. 7A, anchr, lanes 0.48–300]. No apparent effect was observed at these concentrations with the cntrl oligonucleotide [Fig. 7A, cntrl, lanes 0.48–300]. To ensure that the anchr oligonucleotide was inhibiting A complex assembly because of base-pairing with sequences in the anchoring site, we carried out the same experiment using pre-mRNA “e” [see Fig. 6A], which lacks sequences complementary to the anchr oligonucleotide. As shown in Figure 7B, neither the cntrl oligonucleotide, nor the anchr oligonucleotide inhibit A complex assembly with pre-mRNA labeled at the −15 site or “f” pre-mRNA labeled at the −13 site [see Fig. 6B] were incubated in splicing reactions [25 μl] containing the cntrl or anchr 2'O methyl oligonucleotides (2 μg; no oligonucleotide was added for the f cntrl). Reactions were fractionated by gel filtration and the spliceosomal complex UV cross-linked. After treatment with RNase A or T1, proteins were fractionated on a 9% SDS gel. SAPs are indicated.
mRNA "e". Together, these data indicate that base-pairing of an oligonucleotide to the anchoring site inhibits A complex assembly.

To determine whether the 2'O methyl oligonucleotide might block A complex assembly by interfering with the RNA–protein interactions of SF3a/SF3b, we site-specifically labeled pre-mRNA at the −15 site [see Fig. 2] and incubated it under splicing conditions in the presence of the cntrl or anchr oligonucleotide [Fig. 7C]. Spliceosomal complexes were then isolated by gel filtration, UV cross-linked, and treated with either RNase A or T1. A dramatic decrease in cross-linking of the SF3a/SF3b subunits was observed when the anch oligonucleotide was present in the A complex assembly reaction, but not when the cntrl oligonucleotide was present [Fig. 7C, wt pre-mRNA]. Moreover, no effect of the anchr oligonucleotide on SF3a/SF3b cross-linking was observed with pre-mRNA "f" in which the anchr oligonucleotide is not complementary to the anchoring site [Fig. 7C, f pre-mRNA]. Together, these data show that hybridization of the anch oligonucleotide to the anchoring site disrupts cross-linking of the SF3a/SF3b subunits to pre-mRNA.

Discussion

High fidelity in the splicing reaction almost certainly requires the formation of a vast number of RNA–RNA, RNA–protein, and protein–protein interactions during spliceosome assembly. As a first indication of the complexity of the pre-mRNA–protein interactions, we have found that all three subunits of the essential splicing factor SF3a, as well as three of the four subunits of SF3b, UV cross-link to a 20-nucleotide region upstream of the branchpoint sequence in the spliceosomal complex A [see Fig. 8A]. Mutation of this region of the pre-mRNA, designated the anchoring site, does not affect A complex assembly [i.e., stable U2 snRNP binding]. However, when a 2'O methyl oligonucleotide complementary to the anchoring site is added to the splicing reaction, assembly of the A complex is blocked. Moreover, UV cross-linking of the SF3a/SF3b subunits is blocked by addition of the oligonucleotide. Although it is possible that this is attributable to an indirect effect on A complex assembly, the simplest explanation of the data is that the 2'O methyl oligonucleotide blocks the SF3a/SF3b-binding site on the pre-mRNA, thereby preventing A complex assembly. As the SF3a/SF3b proteins are components of U2 snRNP, our data indicate that these proteins function, at least in part, to anchor U2 snRNP to the pre-mRNA in the A complex. Significantly, in another study (Dominski and Kole 1994), a 2'O methyl oligonucleotide complementary to a sequence corresponding in position to our anchoring site was shown to switch branch-site selection in a β-globin pre-mRNA containing duplicated BPs. All of our experiments were carried out using adenovirus major late [AdML] pre-mRNA. Thus, it is likely that the pre-mRNA–protein interactions at the anchoring site are a general requirement for A complex assembly.

Formation of a three-dimensional structure at the anchoring site

The SF3a/SF3b subunits cross-link to the anchoring site in the apparent 5' to 3' linear order SAPs 49, 114, 61, 145, 62, and 155 [Fig. 8A]. This cross-linking order appears to be the same on a pre-mRNA containing a different sequence at the anchoring site. Although we cannot rule out the possibility that some of the cross-linked proteins reflect nonfunctional interactions or are attributable to artifacts imposed by the cross-linking assay [see Results section for additional discussion], it is not an unreasonable conclusion that a large number of U2 snRNP proteins cross-link to the anchoring site in a specific order.

Consistent with the possibility that RNA–protein interactions are dense at the anchoring site, several studies show that this region is highly resistant to RNase digestion in the A complex [Ruskin and Green 1985; Konarska and Sharp 1986; Bindereif and Green 1987; Kramer 1988, Ruskin et al. 1988]. Moreover, the specific, but sequence-independent, order may arise because the pre-mRNA contacts a surface of U2 snRNP in a particular manner.

In addition to these RNA–protein interactions, protein–protein interactions of the SF3a/SF3b subunits have been identified [indicated by arrows in Fig. 8A; Ben-

![Figure 8](genesdev.cshlp.org)
The amino terminus of SAP 49, which contains two RNA-recognition motifs (RRMs), and a carboxy-terminal region in SAP 145 are essential for the SAP 145–49 interaction (Champion-Arnaud and Reed 1994; R. Feld and R. Reed, unpubl.). The interactions of SAPs 61 and 62 with SAP 114 require two distinct regions on SF3a [SAP 114] and amino-terminal regions of SAPs 61 and 62 (Bennett and Reed 1993; Chiarà et al. 1994; Kramer et al. 1995; M. Bennett and R. Reed, unpubl.).

In contrast to the pre-mRNA–protein interactions of the SF3a/SF3b subunits, which have all been detected in functional spliceosomal complexes, the protein–protein interactions between the subunits have only been detected by Far western analysis and coimmunoprecipitation assays. However, it is possible that some or all of the interactions detected with the purified proteins also occur in functional spliceosomal complexes. Arguing in favor of this possibility is the observation that the protein–protein interactions are strong enough in the functional SF3a/SF3b complexes that the subunits copurify over many chromatographic steps (Brosi et al. 1993a,b). In addition, we have obtained no evidence for free SF3a subunits or SF3b subunits in nuclear extracts (M. Bennett and R. Reed, unpubl.). Assuming that the known protein–protein interactions of the SF3a/SF3b subunits do occur in the A complex, our data on the 5′ to 3′ linear cross-linking order of these proteins along the pre-mRNA (Fig. 8A) indicate that some of the proteins that interact with one another are not colinear on the pre-mRNA. Thus, the pre-mRNA or the proteins must be wrapped around one another to form a distinct threedimensional structure. An example of a possible structure in which the pre-mRNA is wrapped around the proteins is shown in Figure 8B. At least part of the function of this structure may be to anchor U2 snRNP on the pre-mRNA. In addition, it is possible that formation of the structure positions the proteins, snRNAs, and/or pre-mRNA for catalysis. Comparison of proteins that cross-link near or at the branch site at later times in spliceosome assembly indicates that the structure at the anchoring site may undergo a major rearrangement as spliceosome assembly proceeds (MacMillan et al. 1994).

Conservation from yeast to humans

Previous work identified the yeast homologs of SAPs 61, 62, and 114 as PRPs 9, 11, and 21, respectively (Bennett and Reed 1993; Brosi et al. 1993b; Legrain and Chapon 1993; Legrain et al. 1993; Ruby et al. 1993; Chiarà et al. 1994; Kramer et al. 1994, 1995; Wells and Ares 1994). The functional conservation between these yeast and mammalian proteins goes beyond the similarity at the amino acid level. In both cases the proteins form a heterotrimer, are present in U2 snRNP, and are required for A complex assembly. In addition, the same protein–protein interactions identified among the mammalian SF3a subunits also occur in yeast (Legrain and Chapon 1993; Legrain et al. 1993; Bennett and Reed 1993; Chiarà et al. 1994; Kramer et al. 1994, 1995). Our present study, together with that of Wells and co-workers (this issue) provides the first evidence that the multimeric splicing factor SF3b may be similarly conserved from yeast to humans. We found that SAP 145 is ~29% identical to a yeast protein in the GenBank data base, and Wells et al. have shown that this yeast protein is identical to CUS1, a gene isolated as a suppressor of a U2 snRNA mutation. Similar to SF3b, CUS1 is essential for A complex assembly (Wells et al., this issue). Moreover, SAP 145 is a component of U2 snRNP, and CUS1 interacts genetically with U2 snRNA. We find that the region in SAP 145 that interacts with SAP 49 (R. Feld and R. Reed, unpubl.) is conserved in CUS1, suggesting that yeast may also have a homolog to SAP 49. Indeed, comparison of SAP 49 to the GenBank data base revealed an S. cerevisiae gene encoding a protein that is 36% identical to SAP 49 (see Wells et al., this issue). This putative yeast SAP 49 consists primarily of the two RRM s that are present at the amino terminus of SAP 49. These RRM s in SAP 49 are required for protein–protein interactions with SAP 145 (Champion-Arnaud and Reed 1994) and likely mediate the pre-mRNA-binding of SAP 49.

As yet there is no report that the yeast SF3a/SF3b subunits bind directly to pre-mRNA in spliceosomal complexes. However, SAP 49 and its probable yeast homolog contain two RRM s, and both SAP 61/PRP9 and SAP 62/PRP11 contain conserved zinc finger-like motifs that are known to mediate interactions with nucleic acids. Thus, it is likely that the yeast SF3a/SF3b subunits interact directly with pre-mRNA. The observation that the SF3a and SF3b subunits are conserved from yeast to humans is consistent with our conclusion that the RNA–protein interactions of these factors at the anchoring site play a critical role in spliceosome assembly.

Materials and methods

cDNA cloning of SAP 145

SAP 145 was obtained from two-dimensional gels of the A complex and used for peptide sequencing as described (Bennett and Reed 1993). The sequence of peptide 1 is EHELLEQK (Fig. 1A, amino acids 151–159) and peptide 2 is PAPELQGVEVALAP-EELDP (Fig. 1A, amino acids 794–817). Degenerate PCR primers were designed to peptide 2. A probe was obtained by RT–PCR of poly(A) RNA and used to isolate a full-length cDNA from a human λ ZAP II bacteriophage library.

Plasmids

pAdML was described in Michaud and Reed (1993). Mutant premRNAs α–4 (Fig. 6) are identical to pAdML except for the 15 nucleotides indicated in the Figure. The mutants were constructed by ligating oligonucleotides containing the 15 randomized nucleotides into the appropriate sites of pAdML.

UV cross-linking of complexes assembled on site-specifically labeled pre-mRNA

Synthesis of AdML pre-mRNA containing a single 32P label was
performed as described (Moore and Sharp 1992). G residues were chosen for site-specific labeling because transcription initiates most efficiently with this nucleotide. Spliceosomal complex A was assembled by incubating 400 ng of site-specifically labeled pre-mRNA in an 0.5-ml splicing reaction for 12.5 min or for the times indicated (Staknis and Reed 1994). Complexes were isolated by gel filtration and UV cross-linked (Champion-Arnaud and Reed 1994). RNase A (4–10 µg) or RNase T1 (4–10 µg) was added to 150 µl aliquots of the gel filtration fractions and incubated for 30 min at 37°C. Increasing the amount of RNase did not change significantly the cross-linking patterns. Proteins were analyzed by gel filtration or two-dimensional gels.

2’ O-methyl oligonucleotide experiments

The cntrl anch' or 2’ O-methyl oligonucleotides (National Bio- sciences, Inc.) are 14-mers complementary to the underlined sequences shown in Figure 7A. Splicing reactions [25 µl] containing AdML or pre-mRNAs “e” or “i” (20 ng) were incubated under standard splicing conditions with the indicated amounts of oligonucleotides for 2.5 min. After quenching the reactions on ice, heparin-loading dye (2.5 mg/ml heparin) was added, and reactions were fractionated on nondenaturing gels. For the cross-linking experiment (Fig. 7C), conditions for oligonucleotide inhibition of splicing were identical except that reactions were 10-fold larger. Samples were fractionated on gel filtration columns and UV cross-linked as above.

Acknowledgments

We are grateful to Maria Bennett, Patrick Champion-Arnaud, David Staknis, and Changyu Wang for contributions to early phases of this work, and we thank members of our laboratory for useful discussions and comments on the manuscript. This work was supported by an National Institutes of Health grant to R.R. The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References

Barabino, S.M.L., B.J. Blencowe, U. Ryder, B.S. Sproat, and A. Lamond. 1990. Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. Cell 63: 293–302.
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.
Bindereif, A. and M.R. Green. 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. EMBO J. 6: 2415–2424.
Brosi, R., H.P. Haut, and A. Kramer. 1993a. Separation of splicing factor SF3 into two components and purification of SF3a activity. J. Biol. Chem. 268: 17640–17646.
Brosi, R., K. Groning, S.E. Behrens, R. Luftmann, and A. Kramer. 1993b. Interaction of mammalian splicing factor SF3a with U2 snRNP and relationship of its 60-kD subunit to yeast PRP9. Science 262: 102–105.
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.
Chiara, M.D., P. Champion-Arnaud, M. Buvoli, B. Nadal-Gi-
Role of highly conserved U2 snRNP proteins

Ruskin, B., P.D. Zamore, and M.R. Green. 1988. A factor U2AF is required for U2 snRNP binding and splicing complex assembly. Cell 52: 207–219.

Staknis, D., and R. Reed. 1994. Direct interactions between pre-mRNA and six U2 snRNP proteins during spliceosome assembly. Mol. Cell. Biol. 14: 2994–3005.

Wells, S.E. and M. Ares. 1994. Interactions between highly conserved U2 small nuclear RNA structures and Prp5p, Prp9p, Prp11p, and Prp21p proteins are required to ensure integrity of the U2 small nuclear ribonucleoprotein in Saccharomyces cerevisiae. Mol. Cell. Biol. 14: 6337–6349.

Wells, S.E., M. Neville, M. Haynes, J. Wang, H. Igel, and M. Ares Jr. 1996. CUS1, a suppressor of cold sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. Genes & Dev. [this issue].

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.
Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A.

O Gozani, R Feld and R Reed

Genes Dev. 1996, 10: Access the most recent version at doi:10.1101/gad.10.2.233