Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex

Dong Zhang and **Michael Rosbash**[1]

Department of Biology, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454 USA

In the yeast commitment complex and the mammalian E complex, there is an important base-pairing interaction between the 5′ end of U1 snRNA and the conserved 5′ splice site region of pre-mRNA. But no protein contacts between splicing proteins and the pre-mRNA substrate have been defined in or near this region of early splicing complexes. To address this issue, we used 4-thiouridine-substituted 5′ splice site-containing RNAs as substrates and identified eight cross-linked proteins, all of which were identified previously as commitment complex components. The proteins were localized to three domains: the exon, the six nucleotides of the 5′ ss region, and the downstream intron. The results indicate that the 5′ splice site region and environs are dense with protein contacts in the commitment complex and suggest that some of them make important contributions to formation or stability of the U1 snRNP–pre-mRNA complex.

Pre-mRNA splicing, the removal of intervening sequences (introns) from eukaryotic transcripts, proceeds through two chemical steps. The 5′ splice site (5′ ss) is first attacked by the 2′-OH of an adenosine residue within the branchpoint, which generates the two intermediates: a 5′ exon and the lariat intermediate. In the second step, the 3′-OH of the 5′ exon attacks the 3′ splice site (3′ ss), which generates the two reaction products: the mRNA (the spliced exons) and the excised lariat intron.

The chemical steps of splicing take place within a large RNA–protein complex, the spliceosome (Moore et al. 1993). Therefore, spliceosome formation precedes the chemical steps of splicing and includes the ordered assembly of numerous protein and small nuclear ribonuclear protein (snRNP) components with the pre-mRNA substrate. In both yeast and mammalian systems, spliceosome formation requires a set of conserved pre-mRNA sequences, largely within the intron. In yeast, these are the 5′ ss (GUAUGU), the branchpoint (UACUA<u>A</u>C), and the 3′ ss (YAG) regions. Therefore, these three sequence elements serve not only as the sites of bond cleavage and formation, but they also recruit splicing factors. Those that are known to interact directly with the pre-mRNA do so through RNA–RNA as well as protein–RNA interactions with the three conserved pre-mRNA regions (Moore et al. 1993; Reed 1996; Berglund et al. 1997; Will and Lührmann 1997).

The earliest event in spliceosome assembly is intron recognition, which presumably targets pre-mRNA to the splicing pathway (Rosbash and Séraphin 1991). In vitro, intron recognition results in a complex containing U1-snRNP, pre-mRNA, and other non-snRNP factors: the E complex in mammals and the commitment complex (CC) in the yeast system (Séraphin and Rosbash 1989; Michaud and Reed 1991). As resolved by native gel electrophoresis, two commitment complexes can be visualized: CC1 and CC2 (Séraphin and Rosbash 1989). Neither complex requires the 3′ splice site, which is also unnecessary for the first chemical step of yeast splicing (Rymond et al. 1987). CC2 formation is dependent on the 5′ ss and the branchpoint regions, whereas CC1 formation only requires the 5′ ss region. Although not certain, it is likely that CC1 is a precursor to CC2 and therefore, the earliest splicing complex. This suggests that the 5′ ss region is the only pre-mRNA-specific element required for initial intron recognition in the yeast system (Rosbash and Séraphin 1991).

Base-pairing between the 5′ ss region and U1 snRNA makes an important contribution to the early splicing complexes in both systems. But this only consists of 5–7 bp, which is probably insufficient to account for the stability of the complexes (Zhuang and Weiner 1986; Séraphin et al. 1988; Siliciano and Guthrie 1988). The yeast commitment complex is especially stable (Legrain et al. 1988; Puig et al. 1999). This suggests that there must be important protein–RNA interactions that stabilize the U1 snRNP–pre-mRNA complex, and there is evidence that supports a role for proteins in early 5′ ss rec-

[1]**Corresponding author.**
E-MAIL rosbash@brandeis.edu; FAX (781) 736-3164.

ognition (Heinrichs et al. 1990; Kohtz et al. 1994; Will and Lührmann 1997). In addition, the cap-binding complex (CBC) is the only identified non-snRNP factor that collaborates with U1 snRNP to help form or stabilize CC1; CBC binds to the cap and also interacts with U1 snRNP through as yet unspecified protein–protein interactions (Colot et al. 1996; Lewis et al. 1996). But there are no other proteins known to interact directly with yeast pre-mRNA substrate to aid the 5′ss–U1 snRNA interaction.

In the yeast system, there has been dramatic recent progress in identifying most and perhaps all of the remaining U1 snRNP protein components (Gottschalk et al. 1998). To date, there are 16 U1 snRNP proteins: 7 Sm proteins and 9 U1 snRNP-specific proteins. The seven Sm proteins are shared with U2, U5, and U4/U6 snRNP. They are well-conserved with mammalian Sm proteins and are proposed to form a seven-member donut structure, which binds to the conserved Sm site present on these snRNAs (Kambach et al. 1999). Of the nine U1 snRNP-specific proteins, three have mammalian orthologs: U1–A, U1–C, and U1–70K. Therefore, it is assumed that their functions are also conserved between yeast and metazoans. Yeast U1–C and U1–70K are essential (or nearly so), but U1–A is not essential (Smith and Barrell 1991; Liao et al. 1993; Tang et al. 1997). U1–A and U1–70K bind directly to yeast U1 snRNA and in very similar locations on the snRNA in yeast and mammals (Kao and Siliciano 1992; Tang and Rosbash 1996), but there is no evidence that U1–C interacts directly with U1 snRNA. Mammalian U1–C may be recruited to U1 snRNP through protein–protein interactions with U1–70K (Nelissen et al. 1994). There are also direct protein–protein contacts between U1–70K and the common D2 and B/B′ proteins, as well as direct contacts between U1–C and B′/B in the mammalian system (Nelissen et al. 1994). It is likely, but not certain, that these interactions also exist in yeast. Finally, there is evidence that U1–C affects the U1 snRNA–5′ ss RNA–RNA interaction (Heinrichs et al. 1990), although there is no evidence that this is through direct U1–C–pre-mRNA contact.

There is little information about the possible functions of the six yeast-specific U1 snRNP proteins: Prp39p, Prp40p, Prp42p (Snu65p), Snu56p, Snu71p, and Nam8p. It is even unknown whether any of them have mammalian orthologs or counterparts. Their locations on the larger yeast U1 snRNA or relative to the 10 universal proteins are also unknown. Prp40p has been proposed to interact with the branchpoint-binding protein (yBBP) (Abovich and Rosbash 1997) during CC2 formation, but has no known role in CC1 formation.

To address possible direct contacts between splicing proteins and the 5′ ss region of a yeast pre-mRNA substrate, we took advantage of the sensitive protein–RNA cross-linking afforded by 4-thiouridine-(4-thioU) substituted RNA. We incorporated 4-thioU at all uridine positions during in vitro RNA synthesis with T7 polymerase. This allowed use of a capped, wild-type rp51A 72 nucleotide 5′ ss-containing RNA for most of the initial experiments. We verified that U1 snRNP complex formation

(CC1 formation) of this RNA is comparable to that of a longer, spliceable substrate. Surprisingly, we detected eight proteins that cross-link to the 5′ ss region. All eight were previously identified splicing proteins: yCBP80, the large subunit of CBC, and seven previously identified U1 snRNP proteins. Using different U-containing substrates and site-specific labeling methods, we localized all eight proteins to three domains: the exon, the intron, and the six nucleotides of the 5′ ss region. The results indicate that the 5′ ss region and environs are dense with direct protein–RNA contacts and suggest that some of them make important contributions to the formation or stability of the U1 snRNP–pre-mRNA complex.
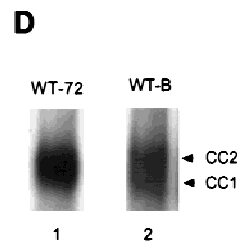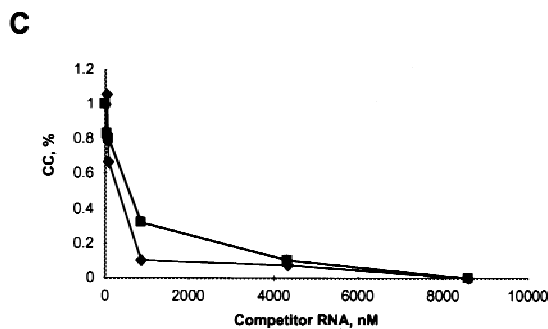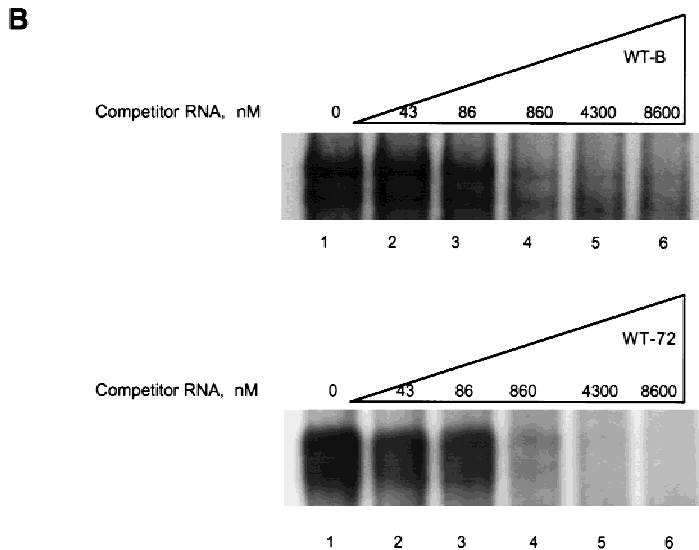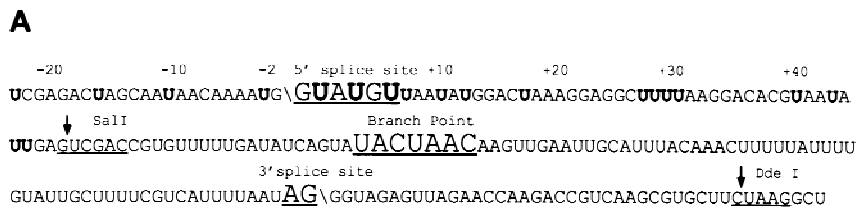
Results

WT-72, a short RNA containing a 5′ss, is a potent CC substrate

To study RNA–protein interactions within or near the 5′ ss region, we used WT-72, an in vitro-synthesized short RNA containing a 5′ ss and a cap (see Materials and Methods). WT-72 contains the first 72 nucleotides of rp51A pre-mRNA, which includes the entire 5′ exon and the first 49 nucleotides of the intron, up to the SalI site (Fig. 1A). A competition assay was used to compare WT-72 with the parental fully competent splicing substrate WT-B (Fig. 1A; Séraphin and Rosbash 1991). Radioactive WT-B was mixed with different concentrations of cold RNA, either WT-72 or WT-B, before the addition of U2-depleted splicing extract. Complex products were then assayed by native gel electrophoresis (Fig. 1B), and the percentage of complex plotted against competitor RNA concentration (Fig. 1C).

WT-72 is almost as good a competitor as WT-B, indicating that WT-72 is a good CC substrate. We also assayed WT-72 complexes directly and compared them with those formed with WT-B. With the WT-B substrate, two complexes were visualized by native gel electrophoresis: CC1 and CC2 (Fig. 1B,D). Previously, it has been shown that both complexes contain U1 snRNP and that only CC2 formation is branchpoint region dependent (Séraphin and Rosbash 1991). WT-72 does not have branchpoint sequence, and only a single complex is formed as expected.

Eight CC proteins are cross-linked in a 5′ ss-dependent manner

To address CC protein–RNA interactions at or near the 5′ ss region, we used 4-thioU-substituted ^{32}P-labeled RNA and cross-linked proteins with long UV irradiation (see Materials and Methods). The WT-72 substrate was incubated in extract, UV irradiated at 365 nm, immunoprecipitated with U1 snRNP-relevant antibody, and the radioactive proteins analyzed by SDS-PAGE. Immunoprecipitation with the 12CA5 monoclonal antibody at 150 mM NaCl, after incubation in a U1–70K–HA-tagged extract (conditions previously used to immunoprecipi-

A

```
        -20             -10           -2  5' splice site  +10        +20        +30        +40
UCGAGACUAGCAAUAACAAAAUG\GUAUGUUAAUAUGGACUAAAGGAGGCUUUUAAGGACACGUAAUA
        ↓   SalI                        Branch Point
UUGAGUCGACCGUGUUUUUGAUAUCAGUAUACUAACAAGUUGAAUUGCAUUUACAAACUUUUUAUUUU
                      3'splice site                                    ↓  Dde I
GUAUUGCUUUUCGUCAUUUUAAUAG\GGUAGAGUUAGAACCAAGACCGUCAAGCGUGCUUCUAAGGCU
```


Figure 1. WT-72 is functional for CC formation. (*A*) Full-length WT-B sequence. WT-B pre-mRNA was transcribed in vitro from the *Dde*I-linearized plasmid pRS195 and is 195 nucleotides in length. WT-72 was transcribed in vitro from the *Sal*I-linearized plasmid and is 72 nucleotides in length. These two restriction sites are indicated (↓). Positions of 4-thioU in WT-72 are in boldface type. The 5′ss, branchpoint, and 3′ splice site regions are in larger font and underlined. \ indicates the splice sites. The 5′ exon sequence has negative numbers, and the intron sequence has positive numbers. (*B*) Competition assay. Radiolabeled WT-B was mixed with increasing amounts of cold competitor RNA, WT-B (*top*), or WT-72 (*bottom*) before the addition of a U2-depleted wild-type extract. After incubation at 25°C for 20 min, the complexes were analyzed by native gel electrophoresis. Nonspecific competitor RNA was without effect. (*C*) Quantification of competition experiment in *B*. Bands corresponding to the CC region of the gel shown in *B* were quantified and plotted against competitor RNA concentration (♦) WT-B; (■) WT-72. Background value was defined as 0% CC, and the CC formed without competitor was defined as 100% CC. (*D*) Native gel analysis of WT-72. Radioactively labeled WT-B (lane *2*) and WT-72 (lane *1*) were incubated in a U2-depleted wild-type Y59 extract at 25°C for 20 min. CCs were analyzed by native gel electrophoresis. Arrows indicate the position of CC1 and CC2.

tate CC), revealed a surprisingly clean pattern of eight proteins (Fig. 2A, lane 1). The 12CA5-generated signal is dependent on a functional 5′ ss as well as an HA tag on U1–70K, but an anti-Nam8p antibody gives rise to the same pattern from a wild-type extract (Fig. 2A, lanes 2–5). There is a completely different pattern of cross-linked proteins in the absence of immunoprecipitation, which is much less obviously affected by the 5′ ss mutant (Fig. 2A, lanes 6,7). Presumably this reflects dominant cross-linking signals from RNP that does not carry U1 snRNP and therefore, is insensitive to the 5′ ss mutation (see below). Surprisingly, the cross-linking pattern was unaffected by addition of the 3′ half of WT-B to WT-72 (Fig. 2B). WT-B was reconstituted by ligating non-thioU and nonradioactive 3′ RNA to WT-72, in which only the first 72 nucleotides were radioactive and thioU substituted. Although we have not assayed the fraction of CC2 formed by this substrate, the implication is that the same proteins associate with the first 72 nucleotides

of CC2 as with CC1. Irradiation at 254 nm indicates that cross-linking at 365 nm is to the 4-thioU (Fig. 2C). Although ~10-fold fainter, the pattern to non-thioU-substituted RNA appears similar. Taken together, the results indicate that the eight proteins associate directly within or near the 5′ ss region of a wild-type rp51A substrate.

YCBP80 and seven U1 snRNP proteins cross-link to the WT-72 pre-mRNA substrate

To identify the 8 proteins, we guessed at candidates based on apparent molecular weight and the 18 known CC1 proteins (CBP80, CBP20, and the 16 U1 snRNP proteins; see introductory section). In the case of candidates encoded by essential genes, we used extracts from strains in which the gene was replaced with a viable partial deletion mutant or functional epitope-tagged version. This should eliminate only the candidate band from the pattern and in favorable cases give rise to a single new band
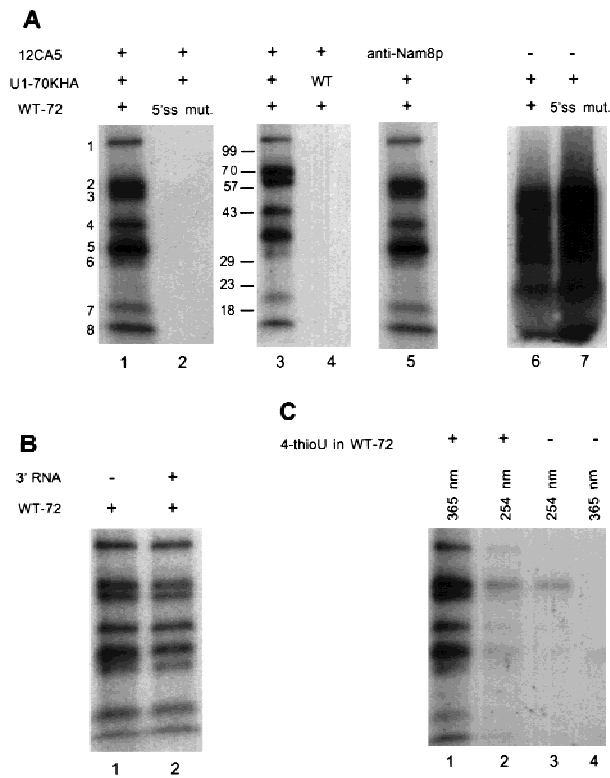
A

B

C

Figure 2. (*A*) Eight proteins cross-link to 4-thioU-substituted WT-72. 4-thioU-substituted ³²P RNAs were transcribed in vitro as described in Materials and Methods. The 5′ ss mutant (5′ ss mut) contains a mutated 5′ splice site (A̲U̲GU̲A̲U̲; mutated nucleotides are underlined). CCs were formed either with an HA-tagged extract (U1–70K–HA) or a nontagged wild-type extract (Y59, WT, lane *4*). After incubation at 25°C for 30 min, complexes were irradiated with 365 nm UV light at 4°C for 5 min. For lanes *1–5*, immunoprecipitation was then done with either anti-HA (12CA5) or anti-Nam8p antibodies as indicated. After immunoprecipitation, beads were treated with RNase T1/A for 20 min at 37°C. Beads were boiled in SDS-polyacrylamide sample buffer, and samples were analyzed on a Tris-HCl 4%–20% linear gradient polyacrylamide gel and autoradiographed. The samples in lanes *6* and *7* were from the same experiment as lanes *1* and *2* but without immunoprecipitation. The extract was treated with RNase T1/A before SDS-PAGE. (*B*) Comparison of the cross-linking pattern of WT-72 with that of WT-B. The substrate in lane *1* is WT-72, as in *A*. The substrate in lane *2* is a ligation product of a 5′ RNA and a 3′ RNA. The 5′ RNA is 4-thioU-substituted ³²P-labeled WT-72. The 3′ RNA is a nonradioactive non-4-thioU-substituted RNA. It was transcribed in vitro using a PCR product as template that carried a T7 promoter and sequences downstream of WT-72 as shown in Fig. 1A. The ligation product was gel purified and used for cross-linking as in *A*. (*C*) Cross-linking patterns of WT-72 substrate irradiated with either 254 nm (lanes *2,3*) or 365 nm UV light (lanes *1,4*). RNA substrates in lanes *3* and *4* were radiolabeled but not 4-thioU substituted.

of altered mobility, as a result of the change in molecular weight. In the case of nonessential genes, we also used extracts from deletion strains (gene knockouts) to verify that the band was absent. To facilitate the use of multiple strains, most immunoprecipitations were done

with anti-Nam8p antibodies (Fig. 3A–D; cf. with Fig. 2A, lane 5).

The molecular weight of the largest protein (protein 1) suggested that it was yCBP80, the large subunit of the CBC. Only this band was absent in extracts from the viable deletion strain yCBP80KO. It was also absent in extract from a viable deletion of its CBC partner protein yCBP20KO (Fig. 3A, lanes 2,3). Previous work indicated that cap binding requires the dimer, consistent with the observation that yCBP80 cross-linking is dependent on both subunits (Izaurralde et al. 1995). The introduction of a plasmid carrying a carboxy-terminal HA-tagged yCBP80 rescues the band (Fig. 3A, lane 5), and higher resolution by 8% SDS-PAGE shows the expected decrease in mobility (Fig. 3A, lane 7 vs. lane 6). We conclude that protein 1 is yCBP80.

Using the same tagged protein–gel mobility strategy, we determined that protein 5 is U1–C; the result is very clean and only one band is altered (Fig. 3B). Similarly, a U1–70K HA-tagged extract indicated that protein 4 is U1–70K (Fig. 3C, lanes 1,2). Although subtle, this mobility change was observed in multiple experiments with multiple extract preparations (data not shown). To verify this conclusion, we also used a U1–70KΔC-1 extract, in which the carboxy-terminal 80 amino acids of U1–70K are absent (Hilleren et al. 1995). Only protein 4 disappeared completely. The entire pattern is faint compared to a wild-type extract, suggesting that CC formation or stability is compromised by the truncated U1–70K protein. But protein 4 is undetectable, consistent with its assignment as U1–70K. (In Fig. 3C, lane 4; note the visible protein 7, which is usually less intense than protein 4.) The absence of a novel band of lower molecular weight might be attributable to comigration with proteins 5–8 or might be attributable to poor cross-linking of the truncated protein to the substrate RNA.

We used HA-tagged proteins to identify the two smallest proteins, 7 and 8 as SmD1 and SmD3, respectively (Roy et al. 1995; Fig. 3D, lanes 1–3). A gift of an SmB–protein A strain from B. Séraphin (EMBL, Heidelberg, Germany) was used to identify protein 6 as SmB (Fig. 3D, lanes 4,5). This is because protein 6 disappeared in the tagged strain. However, no additional lower mobility band was apparent. Although this is probably attributable to poor cross-linking efficiency of the fusion protein, the absence of a new band makes the identification of protein 6 tentative: We cannot exclude an indirect, *trans*-effect of the SmB–protein A on protein 6 cross-linking efficiency. For the sake of simplicity, we will refer to protein 6 as SmB but use an asterisk (SmB*) to indicate the tentative nature of the assignment.

During the course of this work, the Séraphin laboratory reported cross-linking of Nam8p to pre-mRNA within commitment complexes. Without 4-thioU, they observed Nam8p as the most intense cross-linked protein (Puig et al. 1999). This is consistent with protein 2, which also matches the expected molecular weight of Nam8p (see Fig. 2). The use of a Nam8p–protein A strain confirmed this assignment (Fig. 3E, lanes 1,6). As Nam8p is nonessential, we examined a *NAM8* deletion strain.
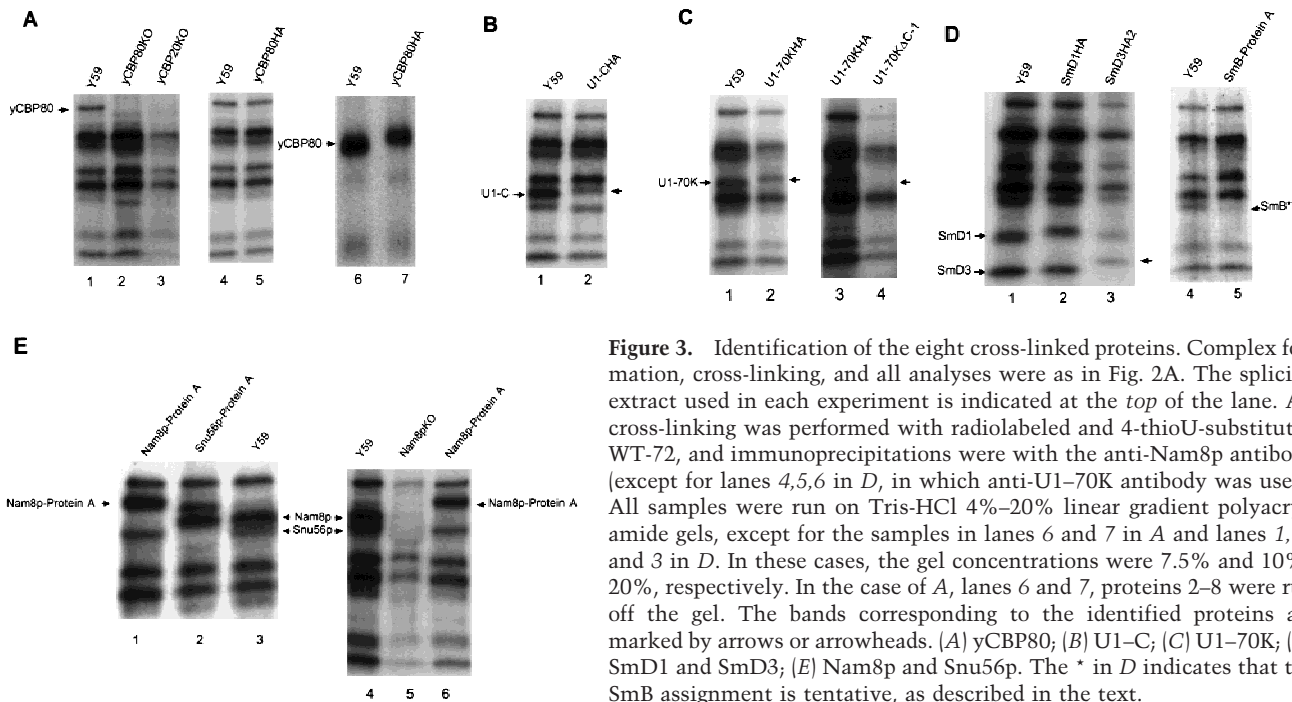
Figure 3. Identification of the eight cross-linked proteins. Complex formation, cross-linking, and all analyses were as in Fig. 2A. The splicing extract used in each experiment is indicated at the *top* of the lane. All cross-linking was performed with radiolabeled and 4-thioU-substituted WT-72, and immunoprecipitations were with the anti-Nam8p antibody (except for lanes *4,5,6* in *D*, in which anti-U1–70K antibody was used). All samples were run on Tris-HCl 4%–20% linear gradient polyacrylamide gels, except for the samples in lanes *6* and *7* in *A* and lanes *1, 2,* and *3* in *D*. In these cases, the gel concentrations were 7.5% and 10%–20%, respectively. In the case of *A*, lanes *6* and *7*, proteins 2–8 were run off the gel. The bands corresponding to the identified proteins are marked by arrows or arrowheads. (*A*) yCBP80; (*B*) U1–C; (*C*) U1–70K; (*D*) SmD1 and SmD3; (*E*) Nam8p and Snu56p. The ★ in *D* indicates that the SmB assignment is tentative, as described in the text.

As predicted, protein 2 is absent from the pattern; however, protein 3, the remaining unassigned band, is also absent. U1 snRNP has been characterized from a *NAM8* deletion strain. In addition to missing Nam8p, the snRNP is also missing Snu56p, another yeast-specific U1 snRNP protein. Because Snu56p matches the molecular weight of protein 3 (Fig. 3E, lane 2), we tested a Snu56p–protein A fusion strain. Protein 3 was apparently converted into a new slower mobility band, confirming the assignment.

Therefore, we have successfully identified all eight bands. Seven are U1 snRNP proteins: two yeast-specific U1 snRNP proteins, Nam8p and Snu56p; two universal U1 snRNP proteins, U1–70K and U1–C; and three Sm proteins, B★, SmD1, and SmD3. One is the large subunit of CBC, yCBP80.

Mapping the binding locations of the eight cross-linked proteins

To provide an initial view of the binding locations of the eight proteins with the same experimental protocol, we used variants of the parental WT-72 substrate. The strategy was to alter the number and positions of U residues, thereby altering the locations of the 4-thioU. Certain variants could not be analyzed, because they decreased substantially CC formation (data not shown). For the same reason, namely, to maintain high levels of complex formation, the three Us at the 5′ ss (intron positions 2, 4, and 6) were not changed.

The first two variants eliminated all exon Us or eliminated all intron Us (Fig. 4A; exon U → A and intron U → A, respectively). These two substrates assigned most of the proteins either to the exon or to the intron (Fig. 4B). The two exceptions are SmB★ and U1–C, which were detectable with both variants. The two proteins must contact the three U's within the 5′ss, or they contact exon and intron Us. Nam8p and Snu56p appeared intron specific. This is consistent with the Nam8p assignment by site-specific labeling in the accompanying paper (Puig et al. 1999). YCBP80, U1–70K, SmD1, and SmD3 appeared exon specific.

To provide a higher resolution map of these four proteins, a small set of exon variants was compared. To maintain uniformly high levels of complex formation, it was necessary to retain the intron Us in this series (Fig. 5A; data not shown); however, this also provided an internal control for overall complex formation: labeling of Nam8p and Snu56p (Fig. 5B). Not unexpectedly, yCBP80 was poorly labeled by the exon–1U variant. It was labeled more intensely as the 4-thioU replacement included more of the 5′ exon, consistent with a predominantly cap-proximal location (Fig. 5B). One of the other proteins, SmD3, was intensely labeled by the exon–1U substrate, consistent with a 5′ss-proximal location. Its labeling increased only marginally with more exon 4-thioUs. U1–70K and SmD1 were intermediate in phenotype; they were clearly labeled by the Exon–1U substrate but increased in intensity with the presence of more exon 4-thioUs. This suggests contact of U1–70K and SmD1 over a greater length of exon, multiple locations in dynamic equilibrium, or multiple subcomplexes (see Discussion). Similar results were obtained in three independent experiments with the same exon substrates, and quantitation relative to Nam8p and Snu56p labeling confirmed these qualitative conclusions (data not shown).
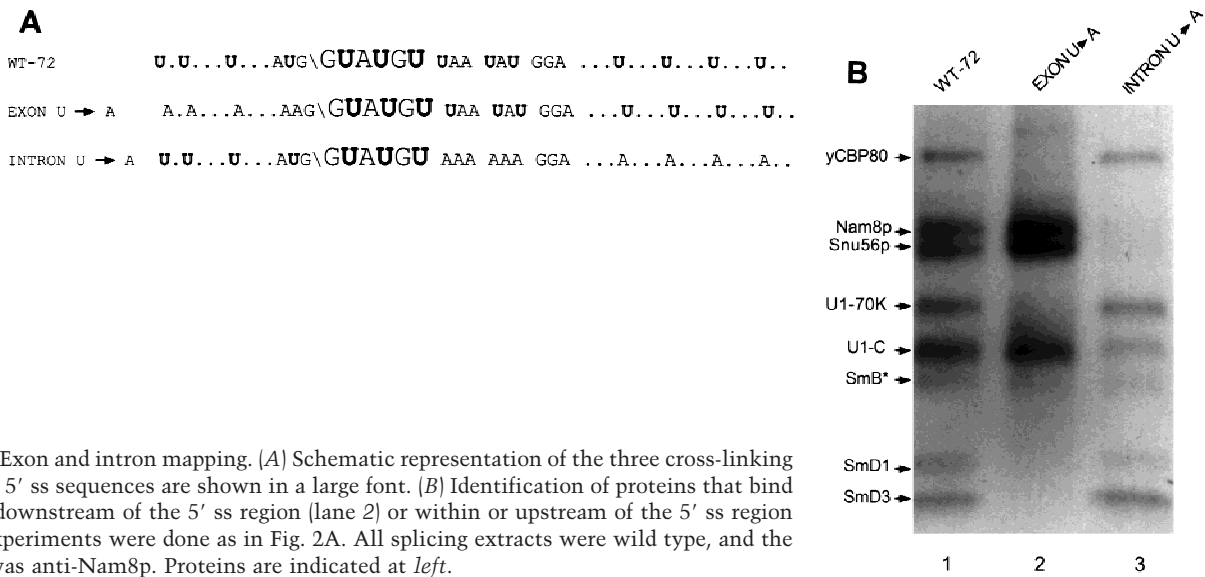
Figure 4. Exon and intron mapping. (*A*) Schematic representation of the three cross-linking substrates. 5′ ss sequences are shown in a large font. (*B*) Identification of proteins that bind within or downstream of the 5′ ss region (lane *2*) or within or upstream of the 5′ ss region (lane *3*). Experiments were done as in Fig. 2A. All splicing extracts were wild type, and the antibody was anti-Nam8p. Proteins are indicated at *left*.

To achieve a preliminary localization of Nam8p and Snu56p, a similar intron mapping strategy was undertaken. For identical reasons to those described above, the exon U's were retained and the exon proteins used as internal controls (Fig. 6). In all cases, Nam8p and Snu56p labeled similarly. Taken together with the genetic and U1 snRNP characterization data, a reasonable working hypothesis is that they contact the intron as a unit. Both proteins label poorly to intron positions 7–12, suggesting predominant contacts to a more distal location. The most intense labeling was achieved by the intron II substrate, placing the region between +18 and +30. But an intermediate level of labeling was obtained with the intron III substrate, suggesting that both proteins also contact the intron between +40 to +46. This region (+18 to +46) is even larger when considering the site-specific location of Nam8p identified in the accompanying paper (+13G). The fact that Nam8p has three RNA-binding domains (RBDs or RRMs) is consistent with an extended

region of contact. Three completely independent experiments and quantitation of two confirmed these qualitative conclusions (data not shown).

Cross-linking to the +2, +4, and +6 U of the 5′ ss region

To address the two proteins with ambiguous labeling profiles (U1–C and SmB*) and to more generally address contacts within the six nucleotides of the 5′ ss region, we used site-specific labeling procedures to generate substrates with a single ^{32}P and a single 4-thioU, either at intron position +2, +4, or +6 (Fig. 7). U1–C was the most strongly labeled protein, especially to +6U but also to +4U. The assignment was confirmed by using a U1–C–HA-tagged extract, which generated the expected mobility decrease. This explains the labeling of U1–C by the exon and intron substrates, because it was substantially cross-linked to the three splice site region Us present in
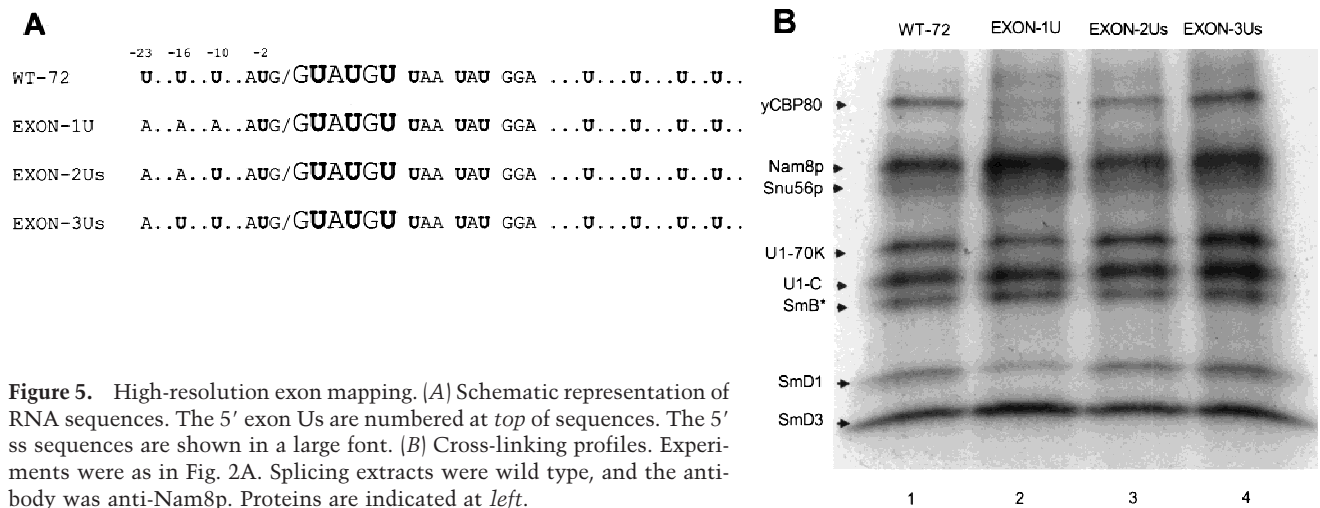

Figure 5. High-resolution exon mapping. (*A*) Schematic representation of RNA sequences. The 5′ exon Us are numbered at *top* of sequences. The 5′ ss sequences are shown in a large font. (*B*) Cross-linking profiles. Experiments were as in Fig. 2A. Splicing extracts were wild type, and the antibody was anti-Nam8p. Proteins are indicated at *left*.
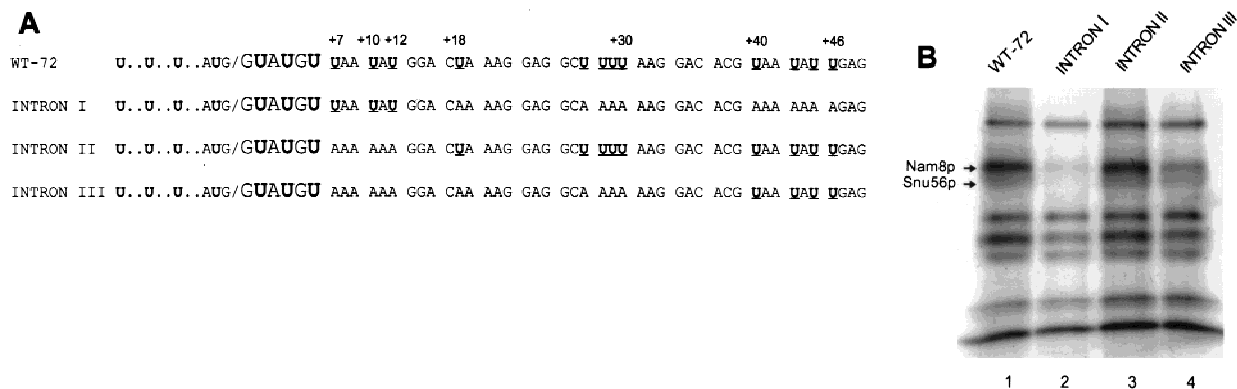
A

| | +7 | +10 +12 | +18 | +30 | +40 | +46 |

WT-72   U..U..U..AUG/G**U**A**U**GU **U**AA **U**A**U** GGA C**U**A AAG GAG GC**U** **UUU** AAG GAC ACG **U**A**U** **U**A**U** **U**GAG

INTRON I   U..U..U..AUG/G**U**A**U**GU **U**AA **U**A**U** GGA CAA AAG GAG GCA AAA AAG GAC ACG AAA AAA AGAG

INTRON II U..U..U..AUG/G**U**A**U**GU AAA AAA GGA C**U**A AAG GAG GC**U** **UUU** AAG GAC ACG **U**AA **U**A**U** **U**GAG

INTRON III U..U..U..AUG/G**U**A**U**GU AAA AAA GGA CAA AAG GAG GCA AAA AAG GAC ACG **U**AA **U**A**U** **U**GAG

B

Nam8p →
Snu56p →

1 2 3 4

Figure 6. High-resolution intron mapping. (*A*) Schematic representation of RNA sequences. Intron Us are numbered at *top* of WT-72 and underlined; 4-thioUs are in boldface type. The 5′ ss sequences are in large font. (*B*) Cross-linking profiles. Experiments were done as in Fig. 2A. The splicing extract was wild type, and the antibody was anti-Nam8p. Nam8p and Snu56p, which are very faint in lane 2 and less prominent in lane 4, are indicated at *left*.

all substrates. SmB* was the second most intensely labeled protein, especially by the +2U substrate and also by the +6U substrate. In the U1–C–HA extract, we note that an increase in SmB* mobility as well as a decrease in U1–C mobility was always observed (e.g., see Figs. 3B

A

| | +2 | +4 | +6 |

+2U ..AUG/G**★U**A UG U AUG..

+4U ..AUG/G UA**★U**G U AUG..

+6U ..AUG/G UA UG**★U** AUG..

B

WT-72   +2U   +4U   +6U

U1-CHA   +   +   +   +

Y59   +   +   +   +

← U1-C

1 2 3 4 5 6 7 8

Figure 7. Cross-linking to the 5′ ss region uridines. (*A*) Schematic representation of the three site-specifically labeled 4-thioU substrates. (★) The position of the single [32]P. Boldface type indicates the adjacent position of the single 4-thioU in each substrate. The 5′ phosphate of the 4-thioU is labeled. (B) Experiments were done as in Fig. 2A. The three RNAs, +2U, +4U, and +6U are ligation products of a 5′ RNA and a 3′ RNA. The 5′ RNAs was a capped, nonradioactive RNA from in vitro transcription. The 3′ RNAs were chemically synthesized with a single 4-thioU at their 5′ ends. The 5′ end of the 3′ RNA was kinased with [α-^{32}P]ATP and ligated to the 5′ RNA (Moore and Query 1998). Corresponding extracts and RNAs are indicated at *top* of each lane . The position of U1-C is marked at *right*.

and 7, lanes 7 and 8). Although there are other possibilities, we interpret the SmB* mobility decrease to indicate that the presence of the tag and the proximity of U1–C and SmB* affect the precise locations of the proteins and ultimately nuclease accessibility. The protein–protein cross-linking of U1–C to SmB in the mammalian system is consistent with this notion (Nelissen et al. 1994).

There are hints of other contacts to the 5′ ss region, but these need to be confirmed by more systematic site-specific labeling experiments. For example, U1–70K may make contact with intron +2U (Fig. 7B, lane 3), although this may represent a very minor contact relative to exon –2U (see Fig. 4). In contrast, the 5′ ss region contains the major sites of U1–C cross-linking. It also appears to contain the major sites of SmB*, but we cannot exclude additional contacts at exon position –2U. Taken together with the exon and intron mapping, the data indicate a surprisingly dense set of protein–RNA contacts within and surrounding the pre-mRNA 5′ ss region–U1 snRNA base-pairing (Fig. 8).
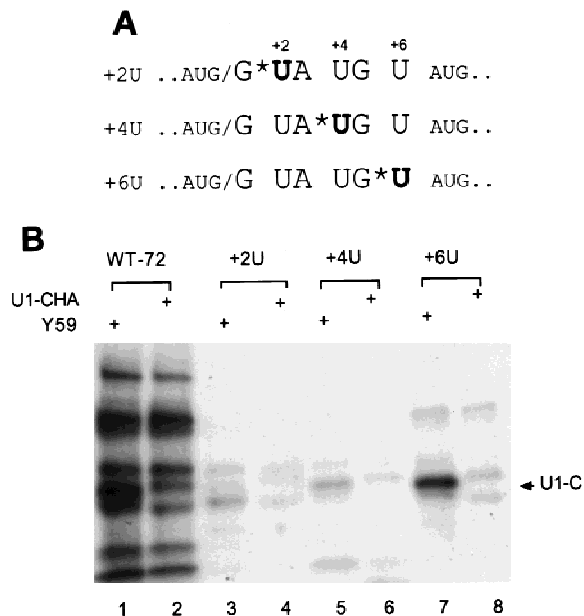
Discussion

In this paper eight proteins are identified that cross-link to the 5′ half of a pre-mRNA substrate within CCs, the product of the first recognized step of in vitro spliceosome formation. The pattern is remarkably clean, and all eight proteins are previously identified CC components: the yCBP 80 and seven U1 snRNP proteins.

While this work was in progress, but before we had identified Nam8p as one of the eight proteins, we learned that the Séraphin laboratory had identified Nam8p in their study (Puig et al. 1999). We confirm here their identification of this protein. The other yeast-specific U1 snRNP protein, Snu56p, appears to map adjacent to Nam8p on the pre-mRNA intron, consistent with genetic experiments suggesting that this pair of proteins might function in a concerted manner (Gottschalk et al. 1998; Fig.7B). As discussed in Puig et al. (1999), it is not yet clear whether mammals have orthologs of these two

Figure 8. Model of the pre-mRNA–protein interactions in the CC1 complex. For U1 snRNA, only the 5' arm, stem/loop I, II, and III are shown. The 5' ss region sequences are shown in lowercase and bold-face letters. The 5' arm sequence of U1 snRNA is shown in uppercase letters. The intron sequences downstream of 5' ss are abbreviated by a thin line; exon sequences are abbreviated by a rectangle and labeled. U1 snRNP-specific proteins are shown in red, Sm proteins in blue, and CBC proteins in green.
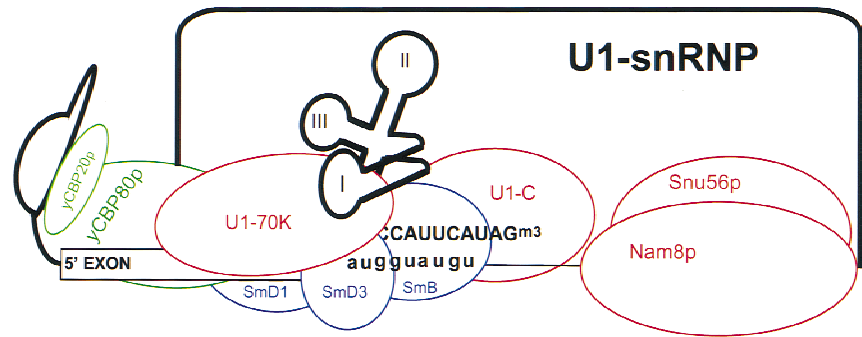
proteins or whether there exist mammalian splicing proteins with a similar functional role.

The other five proteins have well-characterized metazoan counterparts. It will be important to verify that these proteins interact with mammalian pre-mRNA in similar fashion to their yeast counterparts. CBC interacts with U1 snRNP and makes a strong contribution to CC formation (Colot et al. 1996; Lewis et al. 1996). The cross-linking of yCBP80, preferentially near the cap, therefore, is not surprising. It is likely that yCBP20 interacts with the cap through its single RNP domain, and that yCBP80 interacts with the adjacent pre-mRNA (Izaurralde et al. 1994). In the case of U1–C, a direct interaction with pre-mRNA was also anticipated, based on mammalian U1 snRNP–pre-mRNA interaction studies with and without U1–C (Heinrichs et al. 1990; Nelissen et al. 1994). In the case of U1–70K and the three Sm proteins, however, there was no reason to anticipate a positive result.

Recent work on the Sm protein complex provides some indication of how these proteins might bind to the pre-mRNA substrate. Sm D1, D3, and B are the only family members to have carboxy-terminal extensions beyond the Sm motif, suggesting that the residues in these tails might make direct contact with the pre-mRNA substrate (Séraphin 1995). These three proteins are adjacent in the seven-member donut structure, indicating that a single side of the complex contacts the pre-mRNA (Kambach et al. 1999). Therefore, it is tempting to suggest that a function of the Sm complex is to stablilize the U1 snRNP–pre-mRNA interaction, and the Sm-RNA contacts might provide affinity without much specificity (Berglund et al. 1998). It will be interesting to learn whether the same three Sm proteins in another snRNP also contact the pre-mRNA, in which case this function might more generally apply to the Sm complex.

There are probably additional pre-mRNA-binding proteins that cannot form a UV-induced cross-link to the 4-thioUs in this substrate, that is, there are likely to be additional interacting proteins identified by complementary cross-linking approaches or with different pre-mRNA substrates. But it is surprising that there are no nonsplicing-specific candidate proteins in the pattern, that is, hnRNP proteins. For example, Npl3p is known to cross-link to yeast poly(A)$^+$ RNA and has been found associated with U1 snRNP during biochemical purifica-

tion (Wilson et al. 1994; Gottschalk et al. 1998). The absence of this other class of proteins may indicate that CC formation excludes or is incompatible with certain protein–RNA interactions associated with generic RNP. Indeed, the cross-linking pattern in complete extract before U1 snRNP immunoprecipitation is very different, and this complete extract pattern is very similar to that of a mutant splicing substrate (Fig. 2A). In this context, it is surprising that the yCBP80 cross-link is missing or much less apparent in complete extract. As the pattern is clean in this region of the gel, comparable cross-linking of yCBP80 should have been detectable (Fig. 2A). This may indicate that the CBC not only aids the U1 snRNP–substrate interaction but that the CBC–cap interaction is aided or stabilized by U1 snRNP. A predominant role of yeast CBC in splicing is consistent with genetic experiments that focus on CBC and its function (I. Mattaj, pers. comm.).

Our strategy was based on the use of complete splicing extract as well as a 72-nucleotide uniformly substituted 4-thioU substrate, which associates with U1 snRNP in a manner comparable to a full-length wild-type substrate. The use of complete extract rather than purified U1 snRNP was based in part on the existence of important CC factors other than U1 snRNP (e.g., Mud 13p; Colot et al. 1996; Lewis et al. 1996). Indeed, the identification of yCBP80 as a 5' exon-interacting protein suggests that this approach was helpful. The use of a long substrate acknowledged the possibility that all direct interactions might take place within the upstream exon or the downstream intron, despite the interest in identifying proteins associated directly with the pre-mRNA 5'ss–U1 snRNA duplex. This long substrate strategy was also dictated by the notion that a detectable interaction might be an effect as well as a cause of stable, CC formation (i.e., that the highly stable complex formed with a quasi-full-length substrate might aid in protein visualization). Although we are not certain that much shorter substrates do not generate a comparably complex and intense pattern, the location of some proteins quite far from the 5' ss suggests that the long substrate strategy is useful.

The broad scan afforded by the 4-thioU body label is also valuable because it allows subsequent localization of the major sources of signal with no initial bias. Thus, the secondary use of different U-containing substrates provides some indication of where the eight proteins are

positioned on the pre-mRNA substrate. All but two appear to associate exclusively with the exon or exclusively with the intron, and changes of specific Us → As suggests the subregion within which the proteins reside. In the case of Nam8p and Snu56p, this approach placed both proteins somewhat downstream of the position 13 site-specific label used in Puig et al. (1999). Yet the paucity of U residues between positions 12 and 27 makes this conclusion uncertain. The approach also provided a contrast between yCBP80, preferentially positioned near the 5′ end of the 5′ exon, and the Sm proteins, preferentially positioned near or within the 5′ ss region. The results generally suggest that some of these proteins contact a substantial stretch of substrate RNA. This might account for some of the modest sequence bias in 5′ exon sequence (Long et al. 1997), analogous to the intron sequence bias associated with Nam8p binding (Puig et al. 1999). It is likely, however, that many of these contacts take place with only modest sequence specificity, suggesting that the proteins might provide general anchoring functions, analogous to studies on U2 snRNP proteins that bind near the branchpoint region (Gozani et al. 1996; D. Staknis and R. Reed, unpubl.).

Although protein contact over a substantial stretch of substrate RNA is the simplest explanation for the increase in specific signal with increasing 4-thioU content (Fig. 2C), there are alternative interpretations. For example, some proteins may be in dynamic equilibrium between two or more binding locations. Or the observed pattern may reflect the superposition of as yet uncharacterized commitment subcomplexes, within which splicing proteins are fixed at somewhat different substrate locations. But we favor a single static complex, in part because the relative intensity of the individual protein signals is always independent of extract preparation and strain source. Although the entire pattern is faint or strong in different extracts (reflecting weak or potent CC formation), the relative signal from the different proteins is constant (data not shown). Contact along an extended stretch of substrate is likely the case for other RNA binding proteins and adds a note of caution to conclusions drawn exclusively from the site-specific labeling approach; the presence of a protein at a particular location does not preclude its location elsewhere.

Assuming a single CC, we do not know if all of these protein–RNA contacts form simultaneously or if there is a stepwise, ordered assembly pattern. Because six of the eight proteins are U1 snRNP associated, a simple view is that these six contact the substrate in concerted fashion with the 5′ss–U1 snRNA base-pairing. Although the CBC–cap interaction may also form in concert with the U1 snRNP–pre-mRNA interactions, it is attractive to imagine that the former occurs first, at least in vivo. This scenario fits with rapid capping of nascent RNA and suggests that pre-mRNA-associated CBC might even help recruit U1 snRNP before the U1 snRNP–pre-mRNA interaction, as suggested previously (Colot et al. 1996).

It is also not known if all of these contacts are reversed in concerted fashion or stepwise when subsequent steps of spliceosome assembly take place. This is attributable in part to the lack of precise information about whether and when U1 snRNP is removed from more mature splicing complexes. At least some of the 5′ss–U1 snRNA base-pairing must be dissociated before the 5′ss–U6 snRNA interaction, because there are at least two Watson–Crick base-pairs that are identical and therefore, incompatible. These two base-pairs, which occupy the 5′ ss G5 and U6, overlap with the location of U1–C on the 5′ss–U1 snRNA duplex. This suggests that U1–C not only stabilizes the 5′ss–U1 duplex but might also be necessary for the proper addition of U6 snRNP to more mature spliceosome forms. It is even possible that a function of U1–C is to destabilize the 5′ss–U1 duplex, perhaps in concert with U6 snRNP addition.

Although U1–C might stabilize this duplex, the large number of protein–RNA contacts identified in this study provides a different perspective from the traditional view of the pre-mRNA–U1 snRNP interaction. They may provide a much larger energetic contribution to complex stability than the RNA–RNA base-pairing. Although this requires experimental verification, characterization of the CCs without Nam8p (Gottschalk et al. 1998) as well as the salt sensitivity of wild-type CCs (data not shown) is consistent with an important role for proteins. It is even possible that the base-pairing contributes predominantly to complex formation rather than stability, for example, a proofreading or localization function to ensure that the snRNP is correctly positioned before fixation. The fact that there are normally only 5–7 bp suggests furthermore that complex dissociation might occur without extensive helicase activity, as this approximate half-turn of an RNA duplex can disassociate without extensive unwinding. Reversal of the protein–RNA contacts might be much more important in allowing the progression of spliceosome assembly and U6 snRNP addition. The identification of these proteins and their points of contact will hopefully lead to some functional understanding of commitment complex stability as well as mechanisms of complex disassembly during subsequent steps of spliceosome formation.

Materials and methods

Strains and plasmids

For yCBP80-HA, one copy of the 9-amino-acid HA tag was introduced into yCBP80 after the last amino acid by standard procedures, and the coding sequence was cloned into *Bam*HI–*Sal*I site of MC–CUP (Stutz and Rosbash 1994), which generated pDZ46. To construct pDZ32, which contains the DNA template for 5′ ss mutant RNA, two chemically synthesized DNA oligonucleotides were annealed as described below and cloned into *Sac*I–*Xba*I site of pBluescript II (KS+): DZ32a, CTC-GAGACTAGCAATACAAAATGATGTATTATTATGGACT-AAAGGAGGCTTTAAGGACACGTAATATTGAGT (the mutated 5′ splice site is underlined), and its complementary oligonucleotide DZ32b, CTAGACTCAATATTACGTGTCCT-TAAAAGCCTCCTTTAGTCCATATTAATACATCATTTTG-TTATTGCTAGTCTCGAGAGCT. For the construction of the Snu56–protein A plasmid, the promoter and coding region of *SNU56* were recovered by genomic PCR with the oligonucleotides 5′-GATAATACGAAGCAAGAGTCC and 5′-CGATGT-

Table 1. *Strains used in this study*

Strain	Genotype	Reference
Y59	*MAT**a** ade2 arg4 leu2-3,112 trp1-289 ura3-52*	Séraphin et al. (1988)
U1–70KHA	*MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-208 trp1-Δ63 ura3-53 snp1*::pJT–70KHA	this work
ΔyCBP80	*MAT**a** ade2 arg4 leu2-3,112 trp1-289 ura3-52 sto1*::KAN	this work
yCBP80HA	*MAT**a** ade2 arg4 leu2-3,112 trp1-289 ura3-52 sto1*::KAN + pyCBP80HA	this work
ΔyCBP20	*MAT**a** ade2 arg4 leu2-3,112 trp1-289 ura3-52 mud13::ADE2*	Colot et al. (1996)
U1–70KΔC-1	*MAT**a** ade2-1 can1-100 his3 his4 leu2 trp1 ura3 snp1::ADE2* + pΔC-1	Hilleren et al. (1995)
SmD1HA	*MATα leu2-3,112 trp1-289 ura3-52 smd1::LEU2* + pGAL::SMD1HA	Roy et al. (1995)
SmD3HA2	*MAT**a** his1 leu2-1 trp1-Δ101 ura3-52 smd3::LEU2* + pRS314–SMD3HA-3	Roy et al. (1995)
Nam8p–prot. A	*MAT**a** ade2 arg4 leu2-3,112 trp1-289 nam8::URA3* + pNam8p–prot. A	Gottschalk et al. (1998)

TATCCACAGAAAGAGC. After digestion with *Bam*H I and *Sac*I, this fragment was inserted together with a *Hin*dIII–*Sac*I fragment carrying two IgG-binding domain of protein A (Séraphin 1995) into the pRS425 vector (Sikorski and Hieter 1989).

The *SNU56* gene was disrupted as described previously (Gottschalk et al. 1998). A heterozygous diploid was transformed with a *LEU*-marked plasmid carrying a Snu56–protein A fusion protein. After sporulation and tetrad dissection, a viable spore carrying the fusion protein was recovered. Construction of U1–C–HA yeast strains has been described previously (Tang et al. 1997). For other strains used in this paper, see Table 1.

DNA templates for RNA probe

The DNA template for WT-B was linearized at the *Dde*I site of pRS195 (Fig. 1A; Séraphin and Rosbash 1991). The DNA template for WT-72 was linearized at the *Sal*I site of pRS195 (Fig. 1A). The DNA template for the 5′ ss mutant was linearized at the *Xba*I site of pDZ32. The DNA template for exonU → A, intronU → A, exon–1U, exon–2Us, exon–3Us, intron I, intron II, and intron III were prepared as follows. The coding and non-coding DNA strands were chemically synthesized, with the T7 promoter sequence CGCGCGTAATACGACTCACTATA-GGGC (coding strand) at the 5′ end. The two strands were mixed in 50 mM NaCl to make a 0.2 µg/µl solution, heated at 100°C for 3 min, and then transferred to an 80°C heat block. The DNA was cooled slowly by unplugging the heat block and then used directly as a template for in vitro transcription.

Preparation of RNA probes

The RNA probes used in Figures 1 and 2C (lanes 3,4) were labeled radioactively by standard protocols. The preparation of the other RNA probes was similar, with minor modifications. For a 10-µl in vitro transcription reaction, 5 µl of 800 Ci/mmole [α-^{32}P]ATP was dried in a SpeedVac and then mixed with 2 µl of 5× transcription salt [200 mM Tris-HCl (pH 7.9), 50 mM NaCl, 30 mM MgCl$_2$, and 10 mM spermidine], 1 µl of 0.1 M DTT, 1 µl of 10 mM GpppG, 1 µl of NTP mix (0.02 mM ATP, 0.2 mM GTP, 1 mM CTP, and 0.5 mM 4-thio-UTP from USB), 0.4 µl of 40 U/µl RNasin (Promega), 1 µl of 5 mM UTP. For the DNA template from linearized plasmid, 1 µg of DNA and 1 µl of 20 U/µl T7 RNA polymerase (Promega) were added. For the DNA template from annealed synthetic DNA, 1 µl of 0.2 µg/µl DNA and 3.6 µl of 80 U/µl T7 RNA polymerase were added. All RNAs were gel purified.

Preparation of site-specific 4-thioU RNA

All three site-specific 4-thioU RNAs were obtained by ligating two pieces of RNA (Moore and Query 1998). The 3′ RNAs were chemically synthesized; RDZ10b (4-thioUAUGUUAAUAUG-GACUAAAGGAGGCUU), RDZ11b (4-thioUGUUAAUAUG-GACUAAAGGAGGCUU), and RDZ12b (4-thioUUAAUAUG-GACUAAAGGAGGCUU), respectively. They were kinased with [γ-^{32}P]ATP and gel purified. The 5′ RNAs were in vitro transcribed from DNA templates, which were PCR products with *pfu* DNA polymerase (Stratagene) and also gel purified. The DNA bridge for the ligation was AAGCCTCCTTTAGTC-CATATTAACATACCATTTTGTTATTGCTAGTCT, which is complementary to the first 49 nucleotides of WT-72.

Native gel assay and competition analysis

All yeast strains were grown in glucose-containing medium, except that the U1–CHA strain was grown in galactose-containing medium. Preparation of splicing extracts and native gel analysis of splicing complexes were performed as described previously (Abovich et al. 1990). The competition analysis was done as in Séraphin and Rosbash (1991).

UV cross-linking of CC

A 125-µl CC reaction was set up with 70 fmoles of radiolabeled 4-thioU RNA probe (Abovich et al. 1990). After incubation at 25°C for 30 min, an equal volume of Ry buffer [1 µg/µl yeast total RNA, 50 mM HEPES (pH 7.5), 2 mM Mg(OAc)$_2$, and 20 mM EDTA] was added. The complexes were irradiated with 365 nm UV light (model B 100AP, UVP) at 4°C for 5 min.

Immunoprecipitation and gel electrophoresis

Immunoprecipitation was performed essentially as described previously (Banroques and Abelson 1989), with minor modifications. For a 125-µl commitment complex reaction, 10 µl of GammaBind Plus Sepharose bead slurry (Pharmacia Biotech) was washed twice with NET-150 [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.05% NP-40]. The beads were then incubated with antibodies in 150 µl of NET-150 and rotated at 4°C for 1 hr. The beads were washed three times with NET-150 and then incubated with the UV-irradiated reaction mixture in 250 µl of NET-150 at 4°C for 2 hr. The beads were washed three times with NET-150 and then treated with 2 µl of RNase T1/A mix (Ambion) at 37°C for 20 min. After addition of 20 µl of 2× SDS-PAGE loading buffer, the beads were boiled at 100°C for 5 min. The supernatant was analyzed on a Tris-HCl 4%–20% linear gradient polyacrylamide gel and autoradiographed.

References

Abovich, N., P. Legrain, and M. Rosbash. 1990. The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. *Mol. Cell. Biol.* **10:** 6417–6425.

Abovich, N. and M. Rosbash. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89:** 403–412.

Banroques, J. and J.N. Abelson. 1989. PRP4: A protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* **9:** 3710–3719.

Berglund, J.A., K. Chua, N. Abovich, R. Reed, and M. Rosbash. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89:** 781–787.

Berglund, J.A., M. Fleming, and M. Rosbash. 1998. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* **4:** 998–1006.

Colot, H.V., F. Stutz, and M. Rosbash. 1996. The yeast splicing factor MUD13 is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes & Dev.* **10:** 1699–1708.

Gottschalk, A., J. Tang, O. Puig, J. Salgado, G. Neubauer, H.V. Colot, M. Mann, B. Seraphin, M. Rosbash, R. Luhrmann, and P. Fabrizio. 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA* **4:** 374–393.

Gozani, O., R. Feld, and R. Reed. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & Dev.* **10:** 233–243.

Heinrichs, V., M. Bach, G. Winkelmann, and R. Lührmann. 1990. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5′ splice site. *Science* **247:** 69–72.

Hilleren, P.J., H.Y. Kao, and P.G. Siliciano. 1995. The amino-terminal domain of yeast U1-70k is necessary and sufficient for function. *Mol. Cell. Biol.* **15:** 6341–6350.

Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I.W. Mattaj. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78:** 657–668.

Izaurralde, E., J. Lewis, C. Gamberi, A. Jarmolowski, C. McGuigan, and I.W. Mattaj. 1995. A cap-binding protein complex mediating U snRNA export. *Nature* **376:** 709–712.

Kambach, C., S. Walke, R. Young, J.M. Avis, E. de la Fortelle, V.A. Raker, R. Lührmann, J. Li, and K. Nagai. 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*

96: 375–387.

Kao, H.-Y. and P.G. Siliciano. 1992. The yeast homolog of the U1 snRNP protein 70K is encoded by the SNP1 gene. *Nucleic Acids Res.* **20:** 4009–4013.

Kohtz, J.D., S.F. Jamison, C.L. Will, P. Zuo, R. Lührmann, M.A. Garcia-Blanco, and J.L. Manley. 1994. Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* **368:** 119–124.

Legrain, P., B. Séraphin, and M. Rosbash. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* **8:** 3755–3760.

Lewis, J.D., E. Izaurralde, A. Jarmolowski, C. McGuigan, and I.W. Mattaj. 1996. A nuclear cap binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice-site. *Genes & Dev.* **10:** 1683–1698.

Liao, X.C., J. Tang, and M. Rosbash. 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: Implications for snRNP protein function in pre-mRNA splicing. *Genes & Dev.* **7:** 419–428.

Long, M., S.J. de Souza, and W. Gilbert. 1997. The yeast splice site revisited: New exon consensus from genomic analysis. *Cell* **91:** 739–740.

Michaud, S. and R. Reed. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & Dev.* **5:** 2534–2546.

Moore, M.J. and C.C. Query. 1998. Uses of site-specifically modified RNAs constructed by RNA ligation. In *RNA: Protein interactions* (ed. C.W.J. Smith), pp. 75–106. Oxford University Press, Oxford, UK.

Moore, M.J., C.C. Query, and P.A. Sharp. 1993. Splicing of precursors to mRNAs by the spliceosome. In *The RNA world* (ed. R.F. Gesteland and J.F. Atkins), pp. 303–357. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nelissen, R.L., C.L. Will, W.J. van Venrooij, and R. Luhrmann. 1994. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13:** 4113–4125.

Puig, O., A. Gottschalk, P. Fabrizio, and B. Séraphin. 1999. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5′ splice site selection. *Genes & Dev.* (this issue).

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

Rosbash, M. and B. Séraphin. 1991. Who's on first? The U1 snRNP–5′ splice site interaction and splicing. *Trends Biol. Sci.* **16:** 187–190.

Roy, J., K. Kim, J.R. Maddock, J.G. Anthony, and J.L. Woolford Jr. 1995. The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* **1:** 375–390.

Rymond, B.C., D.D. Torrey, and M. Rosbash. 1987. A novel role for the 3′ region of introns in pre-mRNA splicing of *Saccharomyces cerevisiae. Genes & Dev.* **1:** 238–246.

Séraphin, B. 1995. Sm and Sm-like proteins belong to a large family: Identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14:** 2089–2098.

Séraphin, B. and M. Rosbash. 1989. Identification of functional U1 snRNA–pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59:** 349–358.

———. 1991. The yeast branchpoint sequence is not required for the formation of a stable U1 snRNP–pre-mRNA complex and is recognized in the absence of U2 snRNA. *EMBO J.* **10:** 1209–1216.

Séraphin, B., L. Kretzner, and M. Rosbash. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′

cleavage site. *EMBO J.* **7:** 2533–2538.

Sikorski, R.S. and P. Hieter. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae. Genetics* **122:** 19–27.

Siliciano, P.G. and C. Guthrie. 1988. 5′ splice site selection in yeast: Genetic alterations in base-pairing with U1 reveal additional requirements. *Genes* & *Dev.* **2:** 1258–1267.

Smith, V. and B.G. Barrell. 1991. Cloning of a yeast U1 snRNP 70K protein homologue: Functional conservation of an RNA-binding domain between humans and yeast. *EMBO J.* **10:** 2627–2634.

Stutz, F. and M. Rosbash. 1994. A functional interaction between Rev and yeast pre-mRNA is related to splicing complex formation. *EMBO J.* **13:** 4096–4104.

Tang, J. and M. Rosbash. 1996. Characterization of yeast U1 snRNP A protein: Identification of the N-terminal RNA binding domain (RBD) binding site and evidence that the C-terminal RBD functions in splicing. *RNA* **2:** 1058–1070.

Tang, J., N. Abovich, M. Fleming, B. Séraphin, and M. Rosbash. 1997. Identification and characterization of the yeast homologue of U1 snRNP-specific protein C. *EMBO J.* **16:** 4082–4091.

Will, C.L. and R. Luhrmann. 1997. Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9:** 320–328.

Wilson, S.M., K.V. Datar, M.R. Paddy, J.R. Swedlow, and M.S. Swanson. 1994. Characterization of nuclear polyadenylated BNR-binding proteins in *Saccharomyces cerevisiae. J. Cell Biol.* **127:** 1173–1184.

Zhuang, Y. and A.M. Weiner. 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. *Cell* **46:** 827–835.