RNA (1996), 2:1058–1070. Cambridge University Press. Printed in the USA.

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

JIE TANG and MICHAEL ROSBASH

Howard Hughes Medical Institute and Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

ABSTRACT

The yeast U1A protein is a U1 snRNP-specific protein. Like its human counterpart (hU1A), it has two conserved RNA binding domains (RBDs). The N-terminal RBD is quite different from the human protein, and a binding site on yeast U1 snRNA is not readily apparent. The C-terminal RBD is of unknown function. Using in vivo dimethyl sulfate (DMS) protection of mutant strains, we defined a region in yeast U1 snRNA as the likely U1A N-terminal RBD binding site. This was confirmed by direct in vitro binding assays. The site is very different from its vertebrate counterpart, but its location within yeast U1 snRNA suggests a conserved structural relationship to other U1 snRNP components. Genetic studies and sensitive in vivo splicing measurements indicate that the yeast U1A C-terminal RBD also functions in pre-mRNA splicing. We propose that the N-terminal RBD serves to tether the splicing-relevant C-terminal RBD to the snRNP.

Keywords: RNA binding domain; small nuclear RNA; U1 snRNP; splicing

INTRODUCTION

The removal of intervening sequences from pre-mRNA takes place in a large, dynamic nucleoprotein complex called the spliceosome. As a key spliceosomal component, U1 snRNP plays an essential role in pre-mRNA splicing. It is the first snRNP to enter the spliceosome assembly pathway in vitro and undergoes a base pairing interaction with the consensus sequence at the 5′ end of the intron. This pairing contributes to 5′ splice site definition and may commit the pre-mRNA to the splicing pathway (Rosbash & Séraphin, 1991; Lamond, 1993; Madhani & Guthrie, 1994).

In higher eukaryotes, U1 snRNP contains one small nuclear RNA (U1 snRNA) with several well-defined regions that are conserved among different metazoan species (Hamm et al., 1990). In addition to the 5′ arm, which base pairs with 5′ splice sites (Kramer et al., 1984), its stem-loop A and B regions are targets of the U1 70K and U1A proteins, respectively (Query et al., 1989a; Scherly et al., 1989). Its Sm site (a short consensus sequence near the 3′ end) is the principal binding site of the Sm protein complex (Lührmann et al., 1990).

There have been extensive studies in vertebrate systems on the three U1 snRNP-specific proteins (U1C, U1 70K, U1A). U1C is needed for an efficient interaction between U1 snRNP and pre-mRNA 5′ splice sites, but it probably does not bind to U1 snRNA directly (Heinrichs et al., 1990; Nelissen et al., 1991, 1994). U1 70K has an RNA binding domain (RBD) at its N-terminus, which is required for binding to U1 stem-loop A (Query et al., 1989b). At its C-terminus is an arginine–serine-rich region, which may participate in the regulation of spliceosome assembly (Romac & Keene, 1995). The U1A protein has two RBDs, but only the N-terminal RBD is essential for binding to its target, stem-loop B of U1 snRNA (Lutz-Freyermuth et al., 1990). The structure of this RBD has been solved, most recently as a co-crystal with stem-loop B (Nagai et al., 1990; Hoffman et al., 1991; Oubridge et al., 1994). It contains four antiparallel β-sheets and two α-helices connected by loops. The β-sheets form a platform for RNA binding and some of the loops are important for binding specificity (Jessen et al., 1991; Oubridge et al., 1994). In contrast, the function and binding target of the C-terminal

Reprint requests to: Michael Rosbash, Howard Hughes Medical Institute and Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA; e-mail: rosbash@binah.cc.brandeis.edu.

RBD remain elusive. There is no evidence linking this domain to splicing, and it may not even bind to RNA (Lu & Hall, 1995). The only evidence pertaining to function indicates that it interacts with the upstream efficiency element of the SV40 late polyadenylation signal (Lutz & Alwine, 1994).

Because the basic pre-mRNA splicing mechanism is not changed from yeast to human, the general splicing machinery is also substantially conserved. In the yeast *Saccharomyces cerevisiae*, U1 snRNA is five times larger than its metazoan counterpart, yet several of its functionally important regions are conserved, e.g., the 5′ arm, the Sm site, and stem-loop A (Kretzner et al., 1990). Yeast U1 has more than 300 nt of extra sequence in the middle of the molecule, which is conserved in primary and secondary structure among different yeast species (Kretzner et al., 1990). Also, yeast U1 snRNA has a longer stem III, and lacks the well-defined metazoan stem-loop B, the U1A protein binding site.

Due to the low abundance of the yeast U1 snRNP, its protein components have not been fully characterized. However, some evidence suggested that yeast U1 snRNP contains at least seven specific proteins in addition to the common Sm proteins (Fabrizio et al., 1994). To date, four specific yeast U1 snRNP proteins have been cloned (Smith & Barrell, 1991; Liao et al., 1993; Lockhart & Rymond, 1994; Kao & Siliciano, 1996). Two are yeast homologues of human U1 70K and U1A proteins; the others, Prp39p and Prp40p, have no counterpart in human U1 snRNP.

MUD1, the gene encoding the yeast U1 snRNP A protein, was cloned in an enhancer screen (Liao et al., 1993); the *mud1* mutant was lethal in combination with the mutant U1 snRNA used in the screen. Surprisingly, the U1A protein is nonessential in a wild-type U1 snRNA background. Like its human counterpart, it has two RBDs. Both RBDs are conserved between yeast and humans, and the C-terminal RBD shows even higher conservation than the N-terminal one. Epitope-tagged yeast U1A protein can co-immunoprecipitate U1 snRNA from a splicing extract, indicating that it is indeed a U1 snRNP protein (Liao et al., 1993). In this paper, we address two issues: the binding site of yeast U1A on yeast U1 snRNA, and the function of the U1A C-terminal RBD. A combination of in vivo and in vitro approaches identifies the U1A binding site, which is considerably different from its metazoan counterpart. We also provide the first evidence that the C-terminal RBD functions in splicing.

RESULTS

In vivo chemical footprinting of U1 snRNA

Dimethyl sulfate (DMS) enters living cells and methylates nucleic acids (Climie & Friesen, 1988; Ares & Igel, 1990). Total RNA was extracted from the DMS-treated strains, and modification sites were detected by reverse transcription with U1-specific primers. Reverse transcriptase stops before methylated N1-A and N3-C residues (Inoue & Cech, 1985; Moazed et al., 1986). By comparing the DMS modification pattern of the wild-type and Δ*MUD1* strains, we noticed that only two regions on U1 snRNA were modified in the knock out strain, but protected in the wild-type strain (Fig. 1). One is a large loop (loop IIIc) on U1 snRNA stem III, and the other is the single-stranded region opposite to loop IIIc (Fig. 2A). These represent the candidate location for the U1A binding site.

S. cerevisiae U1 snRNA stem III is much larger than its metazoan counterpart. However, loop IIIc and its opposite single-stranded region are reasonable U1A binding-site candidates because their location relative to the rest of the yeast molecule is similar to that of human stem-loop B relative to the rest of the human molecule (Fig. 2B). The yeast loop IIIc is much larger than the human U1 snRNA loop B (21 versus 10 residues), but its size is conserved in another yeast species, *Kluyveromyces lactis* (Kretzner et al., 1990). One striking feature of this loop is its five tandem CA repeats. A similar sequence pattern is found in *K. lactis* U1 snRNA loop IIIc (three CA pairs), but not in human loop B. When part of the loop IIIc sequence was changed to the corresponding *K. lactis* U1 snRNA sequence, or the Cs in the CA repeats were changed to Gs (Fig. 3, KL and GA, respectively), neither the loop nor the single-stranded region opposite to it were protected from DMS methylation (Fig. 4). This indicates that the protection is sequence specific.

To see if the in vivo DMS protection on loop IIIc requires the wild-type yeast U1A protein sequence, we made a series of mutants. The crystal structure of the human U1A N-terminal RBD and U1 stem-loop B complex demonstrated that the four β-sheets interact extensively with RNA, and loop 3 protrudes through the RNA loop (Wharton et al., 1994). Taking this structure as a guide and to minimize structural perturbations, we changed the sequences in the N-terminal RBD β1, β2, and β3 regions into their corresponding human sequences (Fig. 5A: Nβ1, Nβ2, Nβ3). Because the loop 3 sequence is identical between human and yeast U1A (Liao et al., 1993), it was changed to the closely related human U2B″ N-terminal RBD loop 3 sequence (Bentley & Keene, 1991; Fig. 5A: NL3). According to protein secondary structure predictions (Chou and Fasman method), yeast U1A has a much longer loop 1 than human U1A (see Fig. 5 and the Discussion). We made a point mutation that recapitulates the *mud1-4* mutant, which occurs in the predicted loop 1 region (Fig. 5A: NL1; see the Discussion).

When these mutagenized U1A proteins were expressed in the Δ*MUD1* strain, none of them could rescue the DMS phenotype completely (Fig. 6, lanes 2–6). Nβ2, Nβ3, NL1, and NL3 (lanes 3–6) are similar to the

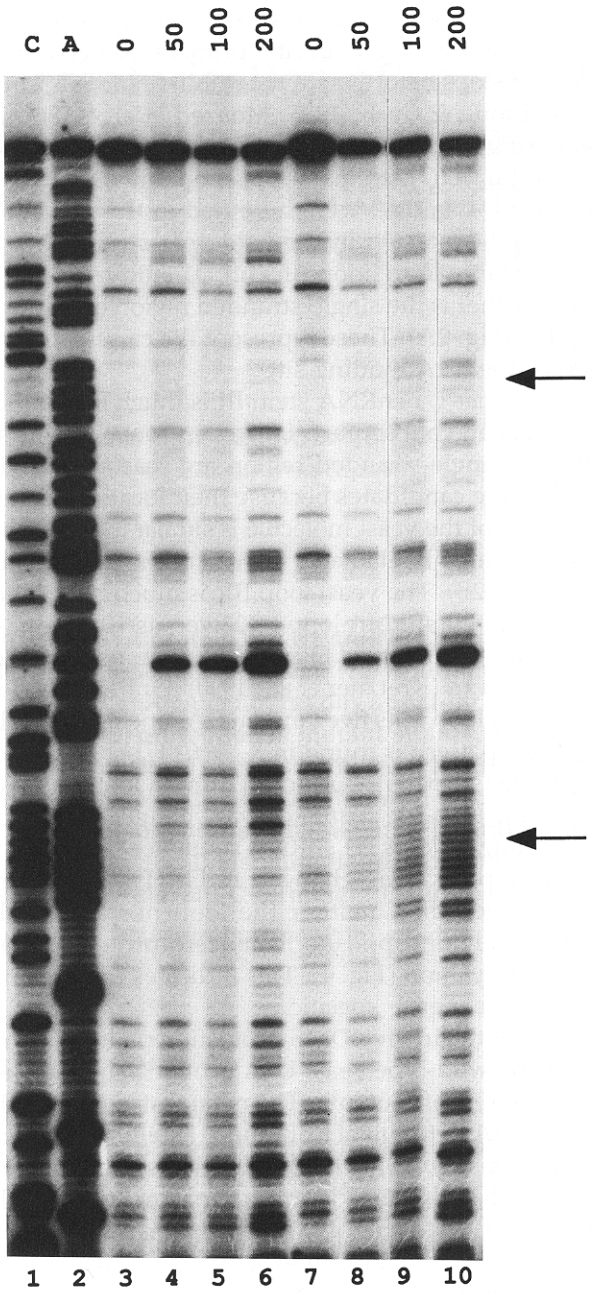

FIGURE 1. In vivo DMS footprinting of U1 snRNA. Wild-type strain (lanes 3–6) or Δ*MUD1* strain (lanes 7–10) grown in 25 mL rich medium was treated with 0 (lanes 3, 7), 50 μL (lanes 4, 8), 100 μL (lanes 5, 9), and 200 μL (lanes 6, 10) of 1:2 diluted DMS. Total yeast RNA was extracted from DMS-treated strains and DMS modification patterns were assayed by reverse transcription with U1-specific primer DT 2502. Extra modification regions for the Δ*MUD1* strain are indicated by arrows. Lower arrow: U1 snRNA loop IIIc; upper arrow: single-stranded region opposite to loop IIIc. Lanes 1, 2, sequence ladder of U1 snRNA.

Δ*MUD1* phenotype (lane 8, vector only), whereas the Nβ1 phenotype (lane 2) is in between that of Δ*MUD1* and wild type (lanes 8 and 9). On the other hand, the C-terminal RBD deletion generates a DMS protection pattern indistinguishable from that of wild-type U1A (Fig. 6, compare lane 7 with lane 9), consistent with the notion that the human U1A C-terminal RBD plays no role in the binding of U1A to its U1 snRNA target (Lutz-Freyermuth et al., 1990).

In vitro nuclease S1 protection

To confirm that U1A protein binds directly to U1 loop IIIc, we performed in vitro binding studies with purified protein and RNA. These assays require relatively large amounts of soluble protein. We overexpressed the U1A N-terminal RBD in several *Escherichia coli* expression systems, but the protein was not soluble (data not shown). We finally succeeded with a thioredoxin fusion system (Novagen), which has been used to express a number of mammalian cytokines and growth factors (LaVallie et al., 1993).

Full-length U1 snRNA was incubated with TrxA-U1A fusion protein and then subjected to partial digestion by nuclease S1. Nuclease S1 specifically cuts single-stranded RNA, and the cleavage sites can be assayed by reverse transcription with U1 snRNA-specific primers. As shown in Figure 7 (upper and lower arrows), loop IIIc and the opposite single-stranded region are protected by the U1A protein, consistent with the in vivo DMS protection data. In addition, loop IIIa (Fig. 7) and loop VIII (data not shown) are also protected from digestion. We suspect that these latter two sites may be due to nonspecific protein binding, because higher protein concentrations were required for the protection of loops IIIa and VIII than for loop IIIc and the single-stranded region opposite to it.

Nitrocellulose filter binding assay

The 122-nt U1 snRNA stem III and its derivatives (see Fig. 3D,E,F) were synthesized in vitro and their binding to purified TrxA-U1A was assayed on nitrocellulose paper. The K_d for full-length U1 snRNA is in the 10^{-9} M range, and the K_d for U1 stem III is even higher. In contrast, binding between the human U1A N-terminal RBD and human U1 stem-loop B is much tighter, with a K_d of 10^{-11} M (Hall & Stump, 1992). But a 3-nt mutation (mut3) or a 6-nt deletion (Δ6) in loop IIIc weakened U1A binding, and a 13-nt deletion (Δ13) almost abolished the binding (Fig. 8A). In a parallel experiment, protein binding was competed with various nonradioactive RNAs (wild-type stem III, mut3, Δ6, and Δ13). As expected, mut3 and Δ6 competed less well with U1A/stem III binding than the wild-type sequence, and Δ13 failed to compete (Fig. 8B). These results confirmed that U1 loop IIIc is a major binding site of yeast U1A.

In vivo splicing assay

A sensitive in vivo assay was applied to compare the effects of the different U1A mutants on pre-mRNA

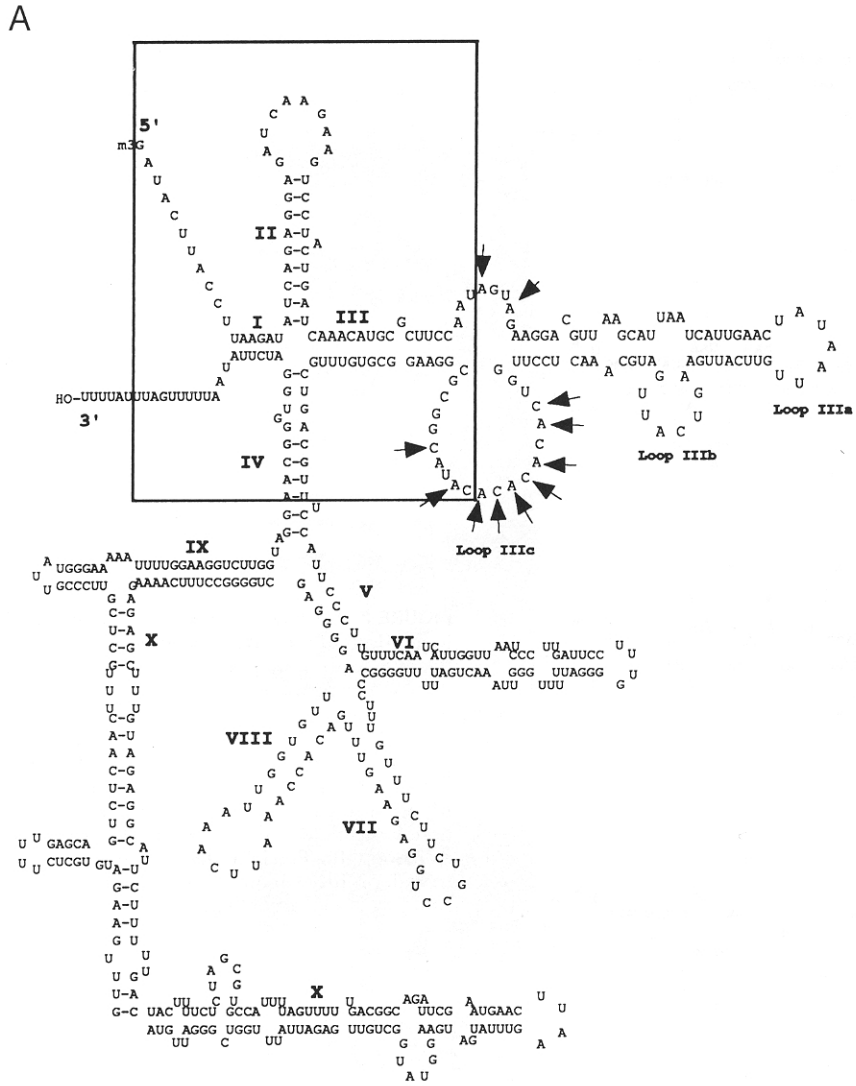

FIGURE 2. Secondary structure comparison of *S. cerevisiae* U1 snRNA and higher eukaryotic U1 snRNA. **A:** *S. cerevisiae* U1 snRNA. The phylogenetically conserved region is shown within a box. DMS modification sites specific for the Δ*MUD1* strain are indicated by arrows. **B:** Higher eukaryotic U1 snRNA. Nonconserved residues are shown as "n"s. The phylogenetically conserved region is shown within a box.

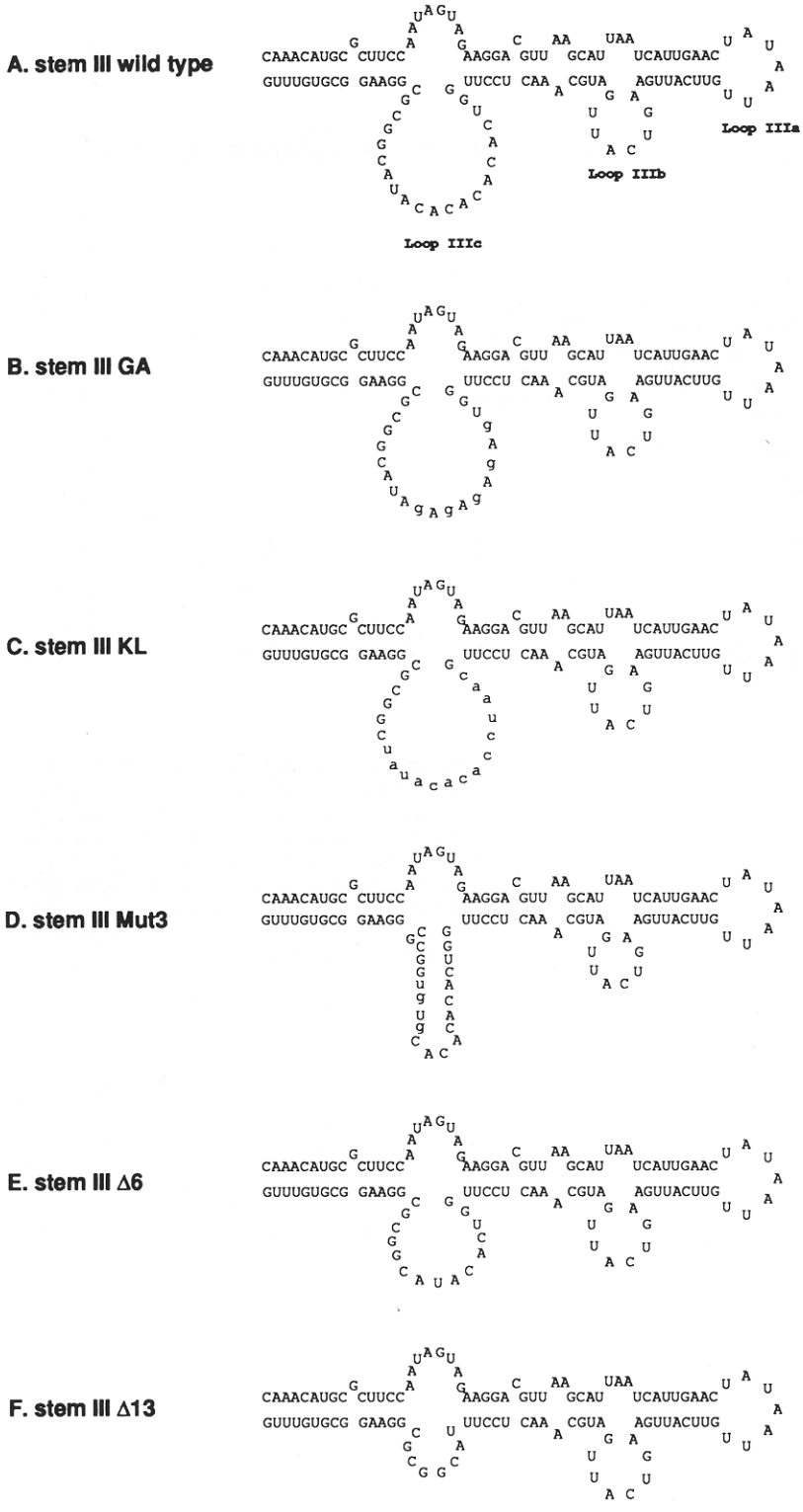

FIGURE 3. U1 snRNA stem III sequence and loop IIIc mutants. **A:** Wild-type stem III sequence. **B:** GA mutant: all the Cs in the five tandem CA repeats are changed to Gs. **C:** KL mutant: part of loop IIIc is changed to corresponding *K. lactis* sequence. **D:** Mut3 mutant: 3 nt at the 3′ end of the loop IIIc are changed, which may form a stem with 5′ part of the loop IIIc according to the computer structure prediction. **E:** Δ6 mutant: three CA repeats are deleted from loop IIIc. **F:** Δ13 mutant: 13 nt are deleted from loop IIIc. Changed nucleotides are indicated by lowercase letters.

splicing. An inefficiently spliced artificial intron sequence (Acc) was inserted into the *CUP1* gene, which encodes a copper chelator protein and allows cells to grow on copper-containing media in a dose-dependent manner (Lesser & Guthrie, 1993; Stutz & Rosbash, 1994; Fig. 9A). With this *CUP1-ACC* reporter, splicing efficiency can be assayed quantitatively by the growth of the strain on different copper concentrations.

In this assay, the Δ*MUD1* strain showed much lower copper resistance (0.2 mM) than the wild-type stain (1.2 mM) (Fig. 9B), consistent with previous observations using an Acc intron-containing β-galactosidase

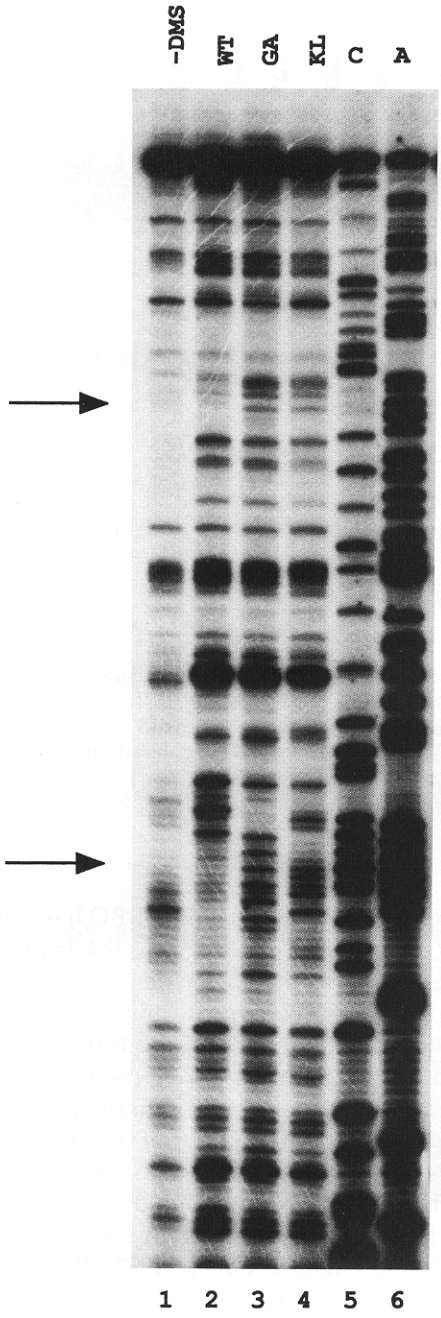

FIGURE 4. DMS protection of U1 snRNA loop IIIc is sensitive to sequence changes in the loop. Wild-type U1 snRNA was replaced by mutant U1 snRNAs by plasmid shuffling in the *mud1-3* strain. Strains carrying KL mutant U1 (lane 4), GA mutant U1 (lane 3), and wild-type U1 (lane 2) were treated with DMS. Total RNA was extracted and assayed by reverse transcription with U1 snRNA specific primer DT 2502. Lanes 5, 6, sequence ladder of wild-type U1 snRNA; lane 1, negative control without DMS treatment. Lower arrow: loop IIIc region; upper arrow: single-stranded region opposite to loop IIIc.

gene as a reporter (Liao et al., 1993). A U1A mutant series was then assayed for its ability to rescue the low copper resistance (i.e., the poor splicing efficiency) of the Δ*MUD1* strain (see Fig. 5). None of the N-terminal RBD mutants fully rescued the phenotype, consistent with their inability to rescue the DMS protection phenotype (Fig. 6). Surprisingly, the C-terminal RBD deletion mutant was indistinguishable from the Δ*MUD1* strain background, i.e., it showed no evidence of biological activity (Fig. 9B), in contrast to its complete rescue of the DMS protection phenotype. Moreover, Cβ1 and Cβ3, two site-specific mutants in the C-terminal RBD (amino acid residues were changed to the corresponding human U1A sequence in an attempt to minimize structural perturbations; see Fig. 5B legend for details) showed a comparable copper resistance to the C-terminal RBD deletion.

We tested several other inefficient introns in the same assay, including a 5′ splice site mutant, a branch point mutant, and interruption of a base pairing region within the intron (Libri et al., 1995). All showed similar splicing defects due to deletion or mutation of the C-terminal RBD (data not shown), suggesting that the C-terminal RBD's contribution to splicing is not limited to certain intron types. Taken together, the data suggest that the C-terminal RBD as well as the N-terminal RBD is important for splicing.

Synthetic lethal between mutant U1 snRNAs and mutant U1A proteins

To test further the in vivo function of the yeast U1A protein, we made several additional alterations to the *MUD1* gene and assayed their genetic interactions with mutant U1 snRNAs. We reported previously that the *mud1-1* allele, which only encodes (the first) 84 amino acids of yeast U1A protein, is synthetic lethal with two U1 snRNA mutants, U1-4U and ΔYC (Liao et al., 1993). Figure 10 reports the growth phenotypes of U1-4U or ΔYC strains transformed with different *mud1* mutants. Three phenotypic classes were observed. First, neither human U1A (the yeast U1A coding sequence replaced by the human sequence) or a H-Y chimera (the yeast N-terminal RBD replaced by the human N-terminal RBD) shows biological activity in this assay; this presumably reflects the limited conservation between human and yeast U1A N-terminal RBDs. Second, the ΔCRBD gene (deletion of the C-terminal RBD) was lethal in combination with U1-4U, but grew slowly with ΔYC. The Y-H chimera (the yeast C-terminal RBD replaced by the human C-terminal RBD) had the same phenotype as ΔCRBD, suggesting that even the more closely related human U1A C-terminal RBD shows no biological activity in yeast. Third, Cβ3 (a 3-amino acid change in the C-terminal RBD β3 region) was lethal with U1-4U, but grew normally with ΔYC. This allele-specific effect could be due to the quantitative difference between the ΔYC and U1-4U mutants, i.e., ΔYC may be a weaker mutant than U1-4U. Alternatively, because the U1-4U mutant has an altered U1 snRNA/5′ splice site base pairing interaction, one can speculate that the U1A C-terminal RBD also contributes, directly

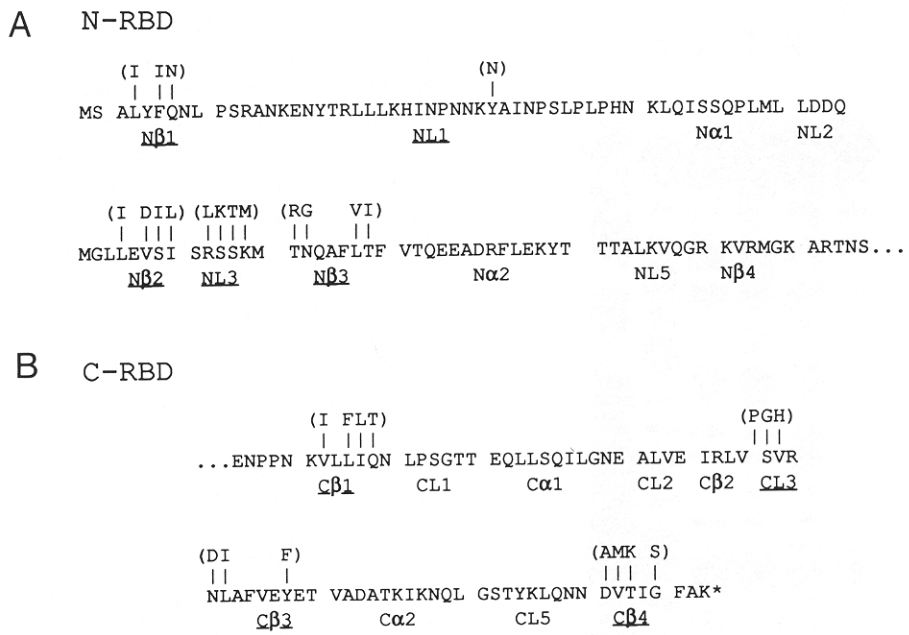

FIGURE 5. Site-specific mutagenesis of yeast U1A protein. **A:** Yeast U1A N-terminal RBD sequence with β-sheets, α-helices, and loops indicated below (mutagenized regions are underlined). Mutations made in Nβ1, NL1, Nβ2, NL3, and Nβ3 regions are shown above the sequence line. **B:** Yeast U1A C-terminal RBD sequence with β-sheets, α-helices, and loops indicated below (mutagenized regions are underlined). Mutations made in Cβ1, CL3, Cβ3, and Cβ4 regions are shown above the sequence line.

or indirectly, to some feature of this base pairing interaction.

DISCUSSION

DMS has been used to probe RNA structure both in vitro (Krol & Carbon, 1989) and in vivo (Ares & Igel, 1990; Zaug & Cech, 1995). However, there has been only limited use of DMS in detecting protein–RNA interactions (Powers et al., 1988) because only a subset of protein–RNA interactions protect the RNA from DMS modification (Muralikrishna & Wickstrom, 1989; Moazed et al., 1995). This is due, in turn, to the fact that the DMS molecule is small so that access to the RNA is often not occluded by protein binding. For example, yeast U1 snRNA loop II was modified by DMS in vivo (data not shown), in spite of the binding of yeast U1 70K protein (Kao & Siliciano, 1992). Presumably, U1 70K protein does not protect the methylation sites on adenines and cytosines of loop II. In our study, most residues in loop IIIc are protected in vivo as well as in splicing extracts (data not shown). We cannot rule out the possibility that a protein complex rather than U1A alone is binding to this loop and protecting it from DMS modification. However, we suspect that there are direct interactions between U1A protein and loop IIIc, based on the in vitro experiments and the observation that DMS protection is sensitive to modest sequence changes in both U1A N-terminal RBD and U1 loop IIIc.

The deletion of the C-terminal RBD has no effect on the DMS protection pattern, i.e., the truncated U1A functions as well as the wild-type protein in protecting loop IIIc from modification. This is consistent with studies on human U1A that the C-terminal RBD is not required for U1 snRNA binding (Scherly et al., 1989; Lutz-Freyermuth et al., 1990). There has been considerable speculation on the roles of proteins with multiple RBDs. In several cases, in vitro binding assays have shown that multiple RBDs worked in *cis* to achieve specific or high-affinity binding to single target RNAs (e.g., U2AF [Zamore & Green, 1991], hnRNP A1 [Shamoo et al., 1994], and sex-lethal [Kanaar et al., 1995]). Based on the studies on yeast and human U1A, we favor an "in *trans*" model for U1A, i.e., the N-terminal RBD tethers the protein to U1 snRNA, whereas the C-terminal RBD touches other RNAs or proteins.

With a soluble thioredoxin fusion protein, we were able to demonstrate in vitro binding to full-length U1 snRNA or U1 stem III. Although several changes in loop IIIc affect yeast U1A N-terminal RBD in vitro binding (Fig. 8), more modest changes have little or no effect, e.g., there was no detectable binding defect when all the Cs in the five tandem CA repeats were changed to Gs (data not shown). Because the GA mutant eliminates in vivo protection, it is possible that the U1A protein binds to its target in two steps. The first is an interaction, perhaps with the RNA backbone, that is not tight enough to prevent DMS modification. The

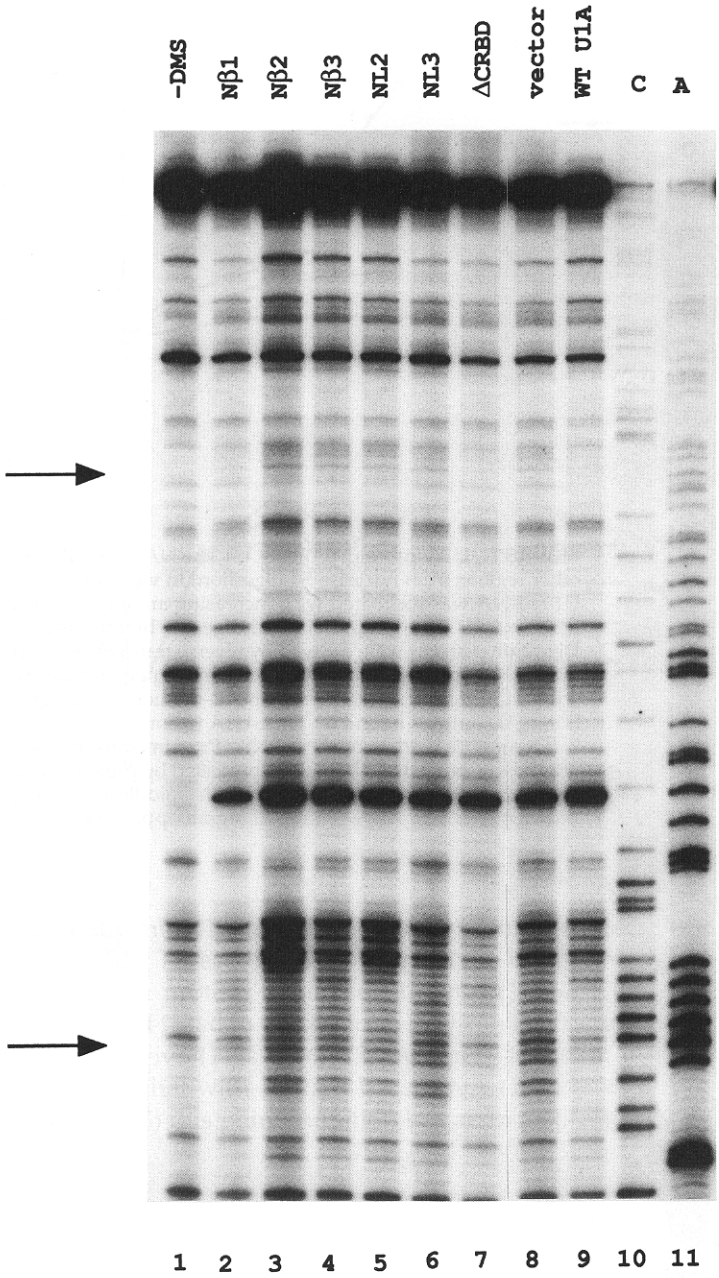

FIGURE 6. DMS protection of U1 snRNA loop IIIc is sensitive to sequence change in the U1A protein. Δ*MUD1* strains carrying different mutagenized U1A proteins were treated with DMS and modification patterns were assayed as described previously. Lane 1, negative control without DMS treatment; lanes 2–6, different site-specific mutants within the N-terminal RBD (as shown above the gel and see Fig. 5 for detail); lane 7, C-terminal RBD deletion; lane 8, vector only; lane 9, wild-type U1A; lanes 10, 11, sequence ladders of U1 snRNA. Lower arrow: loop IIIc region; upper arrow: single-stranded region opposite to loop IIIc.

second is sequence-specific and results in a tight interaction with the loop. The fusion protein or the in vitro-transcribed RNA may not be in an optimal or proper conformation, so the in vitro assays may detect only the first step, which is relatively loose and insensitive to modest sequence changes. (The fusion protein could not protect loop IIIc from DMS modification in vitro; data not shown.) This hypothesis is consistent with recent structural studies on the complex between the human U1A N-terminal RBD and the U1A mRNA 3′ untranslated region, which revealed a two-step recognition mechanism. First, the human U1A N-RBD contacts its RNA target with loop 1 and loop 3, inducing conformational changes in both the protein and RNA. Second, the β-sheets interact extensively with the RNA bases (Allain et al., 1996). This "induced fit" mechanism may also apply to yeast U1A binding. Yet, it is equally likely that additional protein factors contribute to high-affinity U1A binding in vivo, e.g., human U2A′ helps U2B″ bind to its target (Scherly et al., 1990; Boelens et al., 1991).

The combination of the binding and protection results indicates that U1 snRNA loop IIIc is the major binding site of the yeast U1A N-terminal RBD. Its location relative to the 70K binding site, the 5′ arm, and the Sm site is comparable to that of the human U1A binding site. We interpret this as an indication that the relationship of yeast U1A to the rest of the snRNP (i.e.,

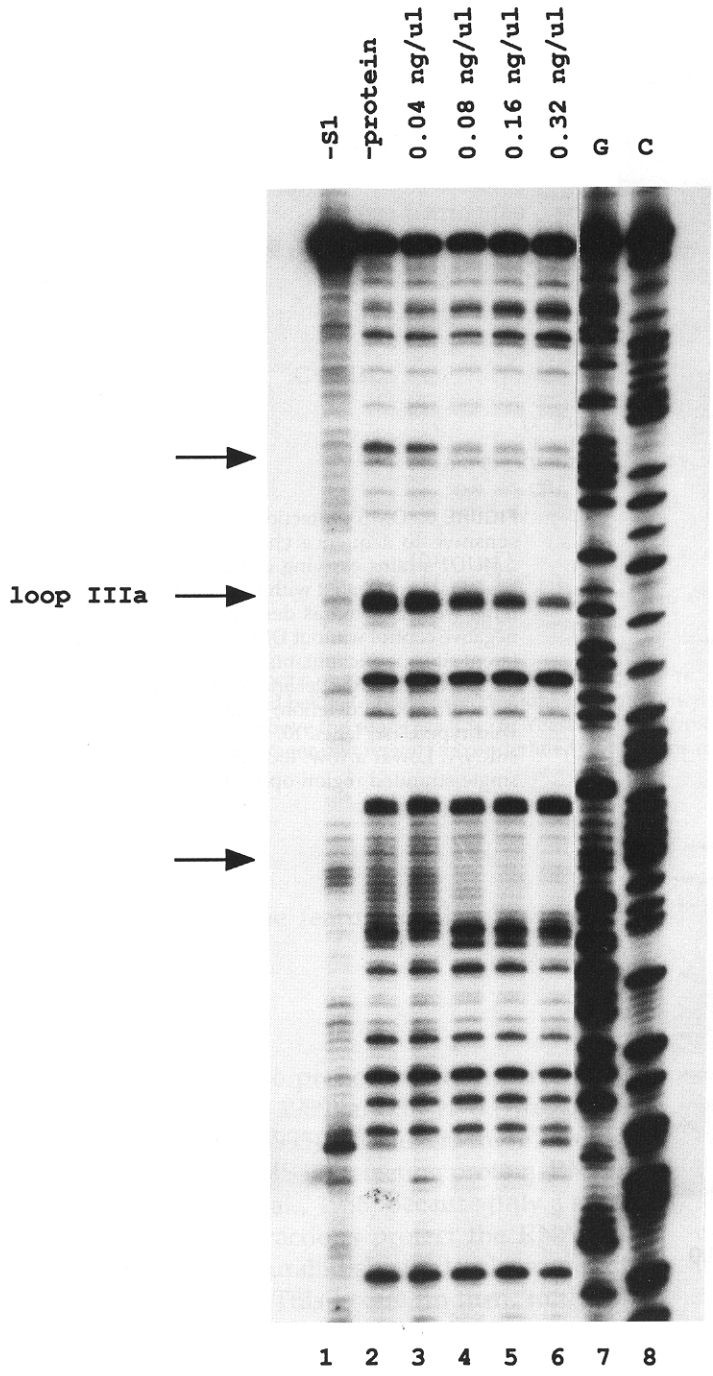

FIGURE 7. U1A protein protects U1 snRNA loop IIIc and other regions from nuclease S1 digestion. In vitro-transcribed U1 snRNA was incubated with increasing amounts of purified TrxA-U1A fusion protein and digested by nuclease S1. Digestion patterns were assayed by reverse transcription with U1 snRNA-specific primer DT 2502. Lane 1, negative control without nuclease S1 treatment; lanes 2–6, nuclease S1 digestion of U1 snRNA bound by TrxA-U1A fusion protein; lanes 7, 8, sequence ladders of U1 snRNA. Protein concentrations are shown above the gel. Regions protected by protein are indicated by arrows. Lower arrow: loop IIIc; middle arrow: loop IIIa; upper arrow: single-stranded region opposite loop IIIc.

the protein–protein interactions it undergoes), as well as its contribution to U1 snRNP function, is conserved between yeast and humans.

Another yeast protein in the U1A-U2B″ protein family, yeast U2B″, has been cloned recently, and its binding site has been defined as yeast U2 snRNA loop IV (Polycarpou-Schwarz et al., 1996; Tang et al., 1996). Yeast U2B″ is highly conserved with human U1A and U2B″, consistent with the phylogenetic conservation among their binding sites, i.e., all three binding sites are well-defined stem-loops with several key conserved nucleotides. In contrast, yeast U1A has an unusual N-terminal RBD with relatively poor conservation with other U1A-U2 B″ family members. A unique feature of this RBD is its extra 34 residues, which we have placed in loop 1, between the β1 and α1 regions. Based on a sequence alignment between human and yeast U1A, these extra sequences were assigned originally to loop 2 (between α1 and β2; Liao et al., 1993). However, secondary structure prediction suggests that the extra sequences are more likely part of loop 1 (between β1 and α1). A much larger loop 1 may coevolve with the

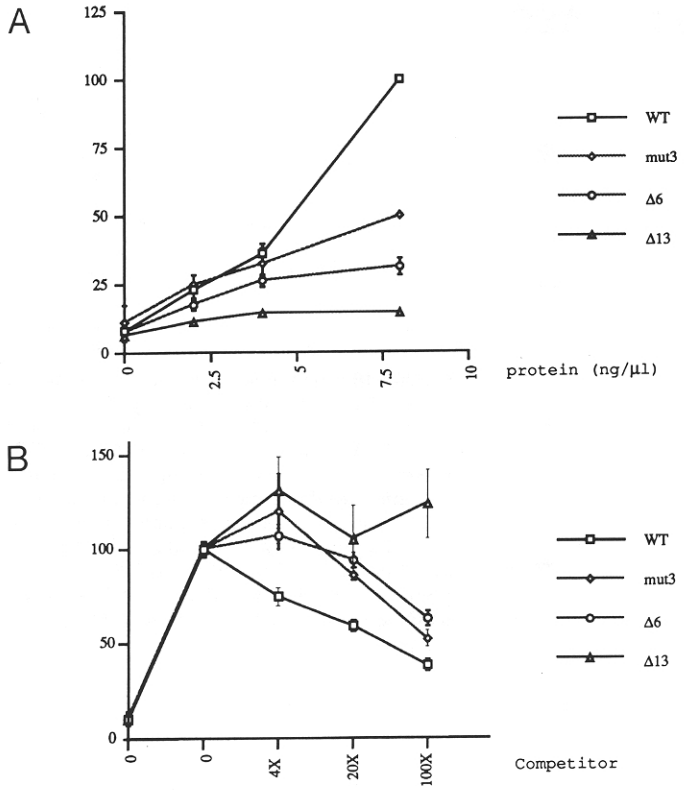

FIGURE 8. U1A protein binds to U1 snRNA stem III specifically. **A:** In vitro-transcribed, ^{32}P-labeled U1 stem III and its mutant versions were incubated with increasing amounts of purified TrxA-U1A fusion protein and bindings was assayed on a nitrocellulose filter. Relative counts retained on the filter were plotted against protein concentration. The highest binding point was arbitrarily set to 100 and the other points were determined accordingly. Two independent experiments were averaged with the bars showing the actual values from these two experiments. At some points, the results from two experiments are so close that the bars may not be seen. **B:** Binding of U1A protein to ^{32}P-labeled U1 snRNA stem III was competed with a 4-, 20-, or 100-fold times excess of cold RNA. Relative counts retained on the filter were plotted against the amount of cold competitor. Two independent experiments were averaged and plotted as in A.

apparently much larger binding site in U1 snRNA, i.e., the loop IIIc. Three arguments support the importance of this extra protein region in binding: (1) the NL1 point mutant failed to protect loop IIIc from DMS modification in vivo (Figs. 5, 6); (2) the *mud1-4* mutant is synthetic lethal with mutant U1 snRNA (Liao et al., 1993); and (3) this point mutation has an in vivo splicing phenotype (Fig. 9). Based on the cocrystal structures of human U1A (Oubridge et al., 1994; Allain et al., 1996), we speculate that the yeast U1A loop 1 may form extensive contacts with loop IIIc to strengthen the binding interaction.

Because *MUD1* is a nonessential gene, it is not surprising that its absence does not cause any detectable

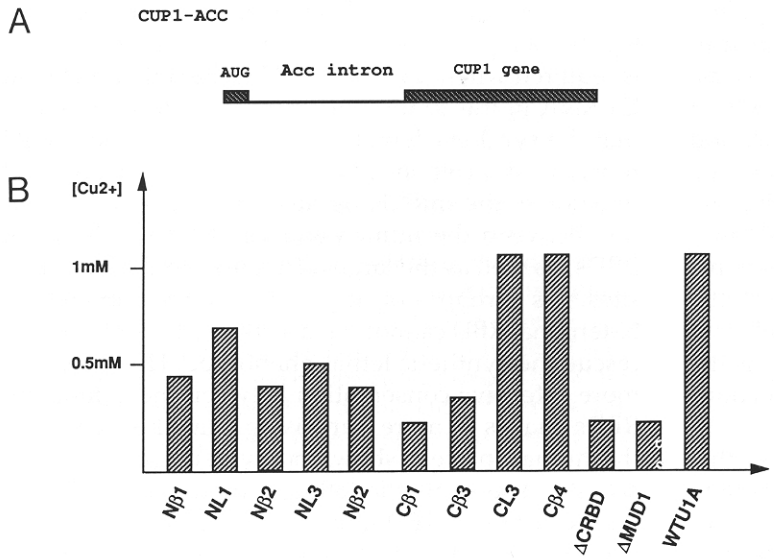

FIGURE 9. In vivo splicing assays. **A:** *CUP1-ACC* reporter gene. *CUP1* gene is interrupted by an artificial intron (Acc) so that only spliced mRNA can encode functional CUP1 product. **B:** Different yeast U1A mutants were introduced into the 72A/Δ*CUP1* strain carrying the *CUP1-ACC* reporter gene. Splicing efficiency of each strain was assayed by its ability to grow on copper-containing medium.

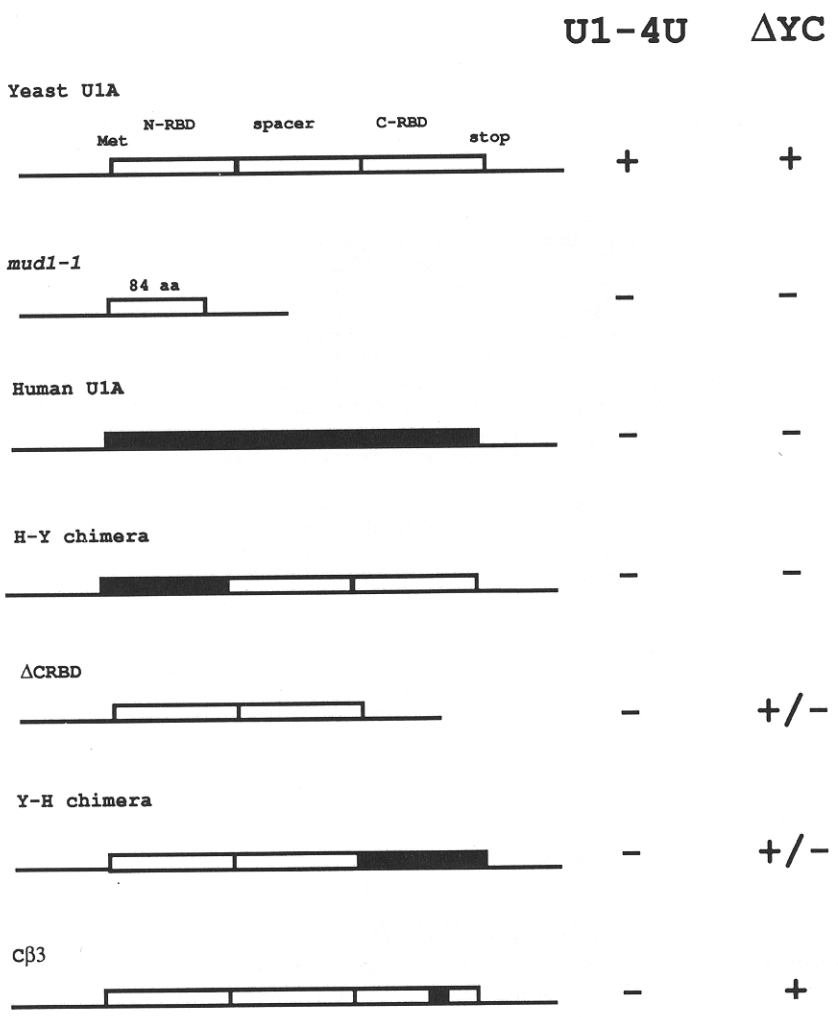

FIGURE 10. Synthetic lethality between mutant U1 snRNAs and mutant U1A proteins. The *mud1-1* host strain carries wild-type and mutant U1 snRNAs (U1-4U or ΔYC) on *URA3* and *TRP1* plasmids, respectively. A series of U1A mutants was introduced into the host strain and synthetic lethality between mutant U1 snRNAs and mutant U1A proteins was assayed on 5-FOA plates, on which the wild-type U1 snRNA was depleted. +, strain lives without wild-type U1 snRNA; −, strain dies; +/-, strain grows slowly.

splicing defect with wild-type introns that have been spliced efficiently. Only with inefficiently spliced introns can we detect a further splicing decrease in the Δ*MUD1* strain. With this assay, it appears that both RBDs are important for splicing. Knowing that all the N-terminal RBD mutants used in this assay failed to protect loop IIIc from DMS modification (Fig. 6), we assume that the effects of N-terminal RBD mutants in splicing reflect their poor binding to U1 snRNA and concomitant effects on U1 snRNP function. On the other hand, deletion of the C-terminal RBD has no effect on binding, so effects of C-terminal RBD mutants on splicing probably reflect a different functional contribution of this domain. We suggest that the N-terminal RBD serves to tether the splicing-relevant C-terminal RBD to the snRNP. In any case, this is the first evidence that the U1A C-terminal RBD functions in pre-mRNA splicing.

ΔYC and U1-4U are the mutant U1 snRNAs that were used for genetic screening and the subsequent cloning of the *MUD1* gene (Liao et al., 1993). ΔYC, with a large deletion in the middle of the U1 molecule, may change the overall conformation of the snRNA and weaken the integrity of the snRNP (Liao et al., 1990). On the other hand, U1-4U is a point mutation within the 5′ arm. It affects base pairing with the 5′ splice site, but may have little effect on U1 snRNP integrity (Séraphin & Rosbash, 1990). Yeast U1A protein is required when ΔYC or U1-4U is the only functional U1 RNA in the strain. An optimistic interpretation is that the synthetic lethal assays with ΔYC and U1-4U reflect U1A's role in, respectively, the structure and function of the snRNP. Because of the poor conservation between the human and yeast U1A N-terminal RBDs, as well as the large differences in their binding sites, it is perhaps not surprising that the human U1A N-terminal RBD cannot replace its yeast counterpart to rescue the synthetic lethal phenotype. However, the more extensive conservation between the C-terminal RBDs makes it more surprising that this region of the human protein shows no evidence of biological activity. The synthetic lethality between U1-4U and U1A C-terminal RBD mutants suggests that the U1A C-terminal RBD may affect the U1 snRNP–pre-mRNA

interaction. Identifying its binding site, either on RNA or on other proteins, is a major goal in further characterization of U1A C-terminal RBD function.

MATERIALS AND METHODS

Strains and plasmids

The *MUD1* knock out strain 72A (MATa, *leu2-112, trp1-289, ura3-52, arg4, ade2,* MUD1D::ADE2) was obtained by replacing the *MUD1* coding sequence with the *ADE2* gene in strain MGD353-13D (Séraphin et al., 1988). The *MUD1* knock out, copper-sensitive strain 72A/ΔCUP1 was derived from strain 72A with the deletion of the endogenous *CUP1* gene as described previously (Stutz & Rosbash, 1994).

All *MUD1* mutants were made by site-specific mutagenesis from plasmid pXL78, which carries the wild-type *MUD1* gene on a *LEU2-CEN* plasmid vector (p366).

All U1 snRNA mutants were made by site-specific mutagenesis from plasmid pXL8 (Liao et al., 1990).

The DNA sequence coding for the first 136 amino acids of the yeast U1A protein was PCR amplified and cloned into the pET-32b vector (Novagen), giving rise to pTrx-U1A plasmid for overexpressing thioredoxin-U1A N-RBD (TrxA-U1A) fusion protein in *E. coli*.

The copper reporter gene *CUP1-ACC* was carried by yeast vector pG-1 (*TRP1,* 2-micron).

In vivo DMS modification

In vivo DMS modifications were done as described previously (Ares & Igel, 1990), except 50, 100, or 200 μL of DMS (diluted 1:2 with ethanol) were added to 25-mL yeast cultures ($OD_{600} \simeq 1$) and incubated for 4 min at 30 °C. RNA was extracted from DMS-treated cells by hot phenol extraction (Ares & Igel, 1990). Ten micrograms of total yeast RNA from each sample were analyzed by reverse transcription with ^{32}P-labeled U1-specific primers and the products were separated by electrophoresis on 5% denaturing polyacrylamide gels. Four U1 specific primers were used to scan 95% of the U1 snRNA molecule and only reactions with primer DT 2502 (nt 266–287) are shown in this paper.

In vitro nuclease S1 protection

TrxA-U1A fusion protein was induced at 25 °C in *E. coli* host strain BL21. The soluble portion of the fusion protein was purified through a nickel affinity column. Full-length yeast U1 snRNA was in vitro transcribed with T7 RNA polymerase and gel purified. In 50 μL S1 buffer (10 mM Tris-HCl, pH 5.5, 10 mM $MgCl_2$, 200 mM KCl, 1 mM $ZnCl_2$, 0.1 μg/μL tRNA, and 0.1 μg/μL BSA), increasing amounts of purified protein were incubated with about 1 pmol U1 snRNA for 15 min at room temperature. Nuclease S1 (5 units) was added subsequently to the reaction and incubated for 15 min. The nuclease S1 digestion was stopped by phenol-chloroform extraction and ethanol precipitation. RNA samples were assayed by reverse transcription with U1-specific primers and the products were analyzed on 5% denaturing polyacrylamide gels. Four U1-specific primers were used to scan 95% of the U1 molecule, and only reactions with DT 2502 are shown in this paper.

Nitrocellulose filter binding assays

U1 snRNA stem III and its mutant versions were in vitro transcribed by T7 RNA polymerase and labeled with ^{32}P-αUTP. Increasing amounts of purified TrxA-U1A protein were incubated with ^{32}P-labeled RNA (10^5 cpm) in 50 μL buffer A (10 mM Tris-HCl, pH 6.8, 2 mM $MgCl_2$, 200 mM KCl, 0.1 mM EDTA, 0.1 μg/μL tRNA, and 0.1 μg/μL BSA) at room temperature for 30 min. The reactions were filtered through nitrocellulose under vacuum and the membrane was washed once with buffer A prior to exposure on a phosphor-imager screen. The signals were quantitated in a Bio-Rad molecular imager system.

The competition assay was done as described above except that TrxA-U1A protein (15 ng/μL) was mixed with 10^5 cpm of ^{32}P-labeled wild-type U1 stem III RNA in the presence of increasing amounts of different cold RNA competitors.

In vivo splicing and synthetic lethal assays

The copper growth assays were performed as described previously (Stutz & Rosbash, 1994). A *MUD1* and *CUP1* double knock out strain (72A/ΔCUP1) was used as host strain.

The synthetic lethal assays were performed as described previously (Liao et al., 1993). The host strain was *mud1-1* with GAL-U1 (on a *URA3* vector) and U1-4U or ΔYC (on a *TRP1* vector).

ACKNOWLEDGMENTS

We thank B. Séraphin, N. Abovich, and H. Colot for constructive comments on the manuscript. This work was supported by the National Institutes of Health grant GM23549 to M.R.

Manuscript accepted without revision August 20, 1996

REFERENCES

Allain FHT, Gubser CC, Howe PWA, Nagai K, Neuhaus D, Varani G. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature 380*:646–650.

Ares MJ, Igel AH. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. *Genes & Dev 4*:2132–2145.

Bentley RC, Keene JD. 1991. Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B″ protein and through interactions with U2 snRNP-A′ protein. *Mol Cell Biol 11*: 1829–1839.

Boelens W, Scherly D, Beijer RP, Jansen EJ, Dathan NA, Mattaj IW, van Venrooij WJ. 1991. A weak interaction between the U2A′ protein and U2 snRNA helps to stabilize their complex with the U2B″ protein. *Nucleic Acids Res 19*:455–460.

Climie SC, Friesen JD. 1988. In vivo and in vitro structural analysis of the rp1J mRNA leader of *Escherichia coli*. Protection by bound L10–L7/L12. *J Biol Chem 263*:15166–15175.

Fabrizio P, Esser S, Kastner B, Luhrmann R. 1994. Isolation of *S. cerevisiae* snRNPs: Comparison of U1 and U4/U6/U5 to their human counterparts. *Science 264*:261–265.

Hall KB, Stump WT. 1992. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res 20*:4283–4290.

Hamm J, Dathan NA, Scherly D, Mattaj IW. 1990. Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J 9*:1237–1244.

Heinrichs V, Bach M, Winkelmann G, Lührmann R. 1990. U1-specific

protein C needed for efficient complex formation of U1 snRNP with a 5′ splice site. *Science 247*:69–72.

Hoffman DW, Query CC, Golden BL, White SW, Keene JD. 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA 88*:2495–2499.

Inoue T, Cech TR. 1985. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: A technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA 82*:648–652.

Jessen T, Oubridge C, Teo CH, Prichard C, Nagai K. 1991. Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J 10*:3447–3456.

Kanaar R, Lee AL, Rudner DZ, Wemmer DE, Rio DC. 1995. Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J 14*:4530–4539.

Kao HY, Siliciano PG. 1992. The yeast homolog of the U1 snRNP protein 70K is encoded by the SNP1 gene. *Nucleic Acids Res 20*(15): 4009–4013.

Kao HY, Siliciano PG. 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol 16*:960–967.

Kramer A, Keller W, Appel B, Lührmann R. 1984. The 5′ terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell 38*:299–307.

Kretzner L, Krol A, Rosbash M. 1990. *Saccharomyces cerevisiae* U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. *Proc Natl Acad Sci USA 87*:851–855.

Krol A, Carbon P. 1989. A guide for probing native small nuclear RNA and ribonucleoprotein structures. *Methods Enzymol*:212–227.

Lamond AI. 1993. The spliceosome. *BioEssays 15*:595–603.

LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. 1993. Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. *BioTechnology 11*: 187–193.

Lesser CF, Guthrie C. 1993. Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, *CUP1*. *Genetics 133*(4):851–863.

Liao X, Kretzner L, Séraphin B, Rosbash M. 1990. Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. *Genes & Dev 4*:1766–1774.

Liao XC, Tang J, Rosbash M. 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.

Libri D, Stutz F, McCarthy T, Rosbash M. 1995. RNA structural patterns and splicing: Molecular basis for an RNA based enhancer. *RNA 1*:425–436.

Lockhart SR, Rymond BC. 1994. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol Cell Biol 14*:3623–3633.

Lu J, Hall KB. 1995. An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J Mol Biol 247*:739–752.

Lührmann R, Kastner B, Bach M. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim Biophys Acta 1087*:265

Lutz CS, Alwine JC. 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes & Dev 8*:576–586.

Lutz-Freyermuth C, Query CC, Keene JD. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc Natl Acad Sci USA 87*:6393–6397.

Madhani HD, Guthrie C. 1994. Dynamic RNA–RNA interactions in the spliceosome. *Annu Rev Genet 28*:1–26.

Moazed D, Samaha RR, Gualerzi C, Noller HF. 1995. Specific protection of 16 S rRNA by translational initiation factors. *J Mol Biol 248*:207–210.

Moazed D, Stern S, Noller HF. 1986. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J Mol Biol 187*:399–416.

Muralikrishna P, Wickstrom E. 1989. *Escherichia coli* initiation factor 3 protein binding to 30S ribosomal subunits alters the accessibility of nucleotides within the conserved central region of 16S rRNA. *Biochemistry 28*:7505–7510.

Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature 348*:515–520.

Nelissen RL, Will CL, van Venrooij WJ, Luhrmann R. 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.

Nelissen RLH, Heinrichs V, Habets WJ, Simons F, Lührmann R, van Venrooij WJ. 1991. Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res 19*:449–454.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.

Polycarpou-Schwarz M, Gunderson SI, Kandels-Lewis S, Séraphin B, Mattaj IW. 1996. *Drosophila* SNF/D25 combines the functions of the two snRNP proteins U1A and U2B″ that are encoded separately in human, potato, and yeast. *RNA 2*:11–23.

Powers T, Stern S, Changchien LM, Noller HF. 1988. Probing the assembly of the 3′ major domain of 16 S rRNA. Interactions involving ribosomal proteins S2, S3, S10, S13 and S14. *J Mol Biol 201*: 697–716.

Query CC, Bentley RC, Keene JD. 1989a. A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol 9*:4872–4881.

Query CC, Bentley RC, Keene JD. 1989b. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell 57*:89–101.

Romac JM, Keene JD. 1995. Overexpression of the arginine-rich carboxy-terminal region of U1 snRNP 70K inhibits both splicing and nucleocytoplasmic transport of mRNA. *Genes & Dev 9*:1400–1410.

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

Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. *Nature 345*:502–506.

Scherly D, Boelens W, van Venrooij WJ, Dathan NA, Hamm J, Mattaj IW. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J 8*:4163–4170.

Séraphin B, Kretzner L, Rosbash M. 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.

Séraphin B, Rosbash M. 1990. Exon mutations uncouple 5′ splice site selection from U1 snRNA pairing. *Cell 63*:619–629.

Shamoo Y, Abdul-Manan N, Patten AM, Crawford JK, Pellegrini MC, Williams KR. 1994. Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry 33*:8272–8281.

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

Tang J, Abovich N, Rosbash M. 1996. Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B″ protein. *Mol Cell Biol 16*:2787–2795.

Wharton KAJ, Franks RG, Kasai Y, Crews ST. 1994. Control of CNS midline transcription by asymmetric E-box-like elements: Similarity to xenobiotic responsive regulation. *Development 120*:3563–3569.

Zamore PD, Green MR. 1991. Biochemical characterization of U2 snRNP auxiliary factor: An essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J 10*:207–214.

Zaug AJ, Cech TR. 1995. Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: Telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA 1*:363–374.