MOLECULAR AND CELLULAR BIOLOGY, June 1996, p. 3012–3022 Vol. 16, No. 6
0270-7306/96/$04.00+0
Copyright © 1996, American Society for Microbiology

# Base Pairing at the 5′ Splice Site with U1 Small Nuclear RNA Promotes Splicing of the Upstream Intron but May Be Dispensable for Splicing of the Downstream Intron

DONG-YOUN HWANG AND JUSTUS B. COHEN*

*Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261*

Received 26 January 1996/Returned for modification 28 February 1996/Accepted 15 March 1996

**We previously reported that exon skipping in vivo due to point mutations in the 5′ splice site (5′ss) signal of an internal mammalian exon can be prevented by coexpression of U1 small nuclear RNAs, termed shift-U1s, with complementarity to sequences upstream or downstream of the mutated site. We now show by S1 nuclease protection experiments that a typical shift-U1 restores splicing of the upstream intron, but not necessarily of the downstream intron. This indicates that the normal 5′ss sequence acts as an enhancer for splicing of the upstream intron, that it owes this activity to base pairing with U1, and that the enhancer activity is reproduced by base pairing of U1 with other sequences in the area. Shift-U1s are dispensable when the 3′ss sequence of the upstream intron is improved, which suggests that base pairing of U1 with sequences at or near the downstream end of the exon normally functions by compensating for a weakness in the upstream 3′ss. Accordingly, U1 appears to be involved in communication across the exon, but our data indicate at the same time that extensive base pairing between U1 and the 5′ss sequence is not necessary for accurate splicing of the downstream intron. These findings are discussed in relation to the coordinate selection of exon termini proposed by the exon definition model.**

The removal of introns from mRNA precursors (pre-mRNA) involves the recognition of sequence signals at the intron's termini, the 5′ splice site (5′, ss) and 3′ss signals, and the region within the intron that contains the branch point and polypyrimidine tract. Recognition results in the assembly of a multicomponent catalytic complex, the spliceosome, through a series of dynamic interactions between four small nuclear ribonucleoprotein particles (snRNPs), an unknown number of proteins, and pre-mRNA. The consequent splicing reaction is a two-step process (for reviews, see references 13, 23, and 25).

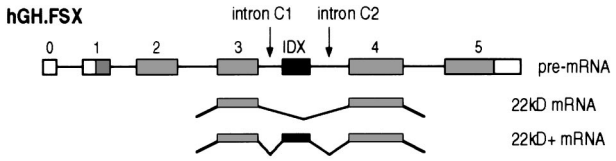
Accuracy in splicing is essential for proper expression of the many eukaryotic genes that contain introns. In relatively simple organisms, such as the yeast *Saccharomyces cerevisiae*, the information contained in the splicing signals may be largely sufficient to direct accurate splicing. The sequences at the 5′ss and branch point are highly conserved, and solitary, relatively short introns appear to be the rule (32, 49). In contrast, the sequence signals in complex organisms such as humans are degenerate, many pre-mRNAs contain multiple introns, and the introns may be very large. While exons must be joined in the correct order, flexibility must be maintained to permit alternative splicing (13, 25).

Several clues have emerged in recent years to explain how the complex demands on the splicing machinery in higher eukaryotes are met. At the most fundamental level, it was proposed that pairs of splice site signals are recognized with reference to one another across exons, not introns (30). This could explain the observation that mutations at either splice site frequently result in exon skipping and implies communication between splice sites across exons (reviewed in reference 2). Indeed, it has been demonstrated that base pairing of the RNA component of the U1 snRNP, U1 snRNA, with the 5′ss signal can stimulate binding of the general splicing factor $U2AF^{65}$ to the polypyrimidine tract of the preceding intron (15); $U2AF^{65}$ is necessary for the recruitment of U2 snRNP to the branch point region (reviewed in references 25 and 31). It was suggested that the inferred interaction between U1 snRNP and $U2AF^{65}$ is indirect and requires additional proteins (15). In vitro experiments support the possibility that members of the SR family of proteins, essential splicing factors characterized by RNA-binding and arginine-serine-rich domains (53), can provide a bridging function (50). This would be consistent with the observation that the SR protein SF2/ASF promotes exon inclusion and diminishes exon skipping (3, 24).

Another recent development is the recognition that the removal of introns with weak splicing signals can be greatly stimulated by specific sequences within exons, usually the flanking exon downstream (14, 16, 26, 48, 52). These sequences, termed splicing enhancers, appear to function as binding sites for proteins that somehow facilitate splicing of the upstream intron (21, 29, 39, 41, 43, 44), perhaps similar to the stimulation of $U2AF^{65}$ binding (referred to above) by U1 across the exon. U1 RNA and U1 proteins have been detected in splicing enhancer-binding complexes (21, 39, 45, 48, 52), and there is some evidence that U1 RNA may pair with exonic enhancer sequences (48). However, it is not clear that these are general conditions for enhancer activity. On the other hand, it has been shown that 5′ss sequences can also act as splicing enhancers and that this requires the involvement of U1 (20, 39, 48).

U1 RNA normally pairs with the 5′ss signal, i.e., the conserved 9-nucleotide (nt) sequence element at exon-intron boundaries (see Fig. 1), and it was initially believed that this interaction alone determines the position of the 5′ss (e.g., reference 1). However, the situation is probably more complex (31), and it now appears that the position of the 5′ss is defined relative to base pairing interactions between pre-mRNA and the U5 and U6 snRNAs (8, 18, 22, 27, 28, 33, 38, 47, 51), while

* Corresponding author. Electronic mail address: jbc+@pitt.edu.




5'ss mutations

| | | | |
|---|---|---|---|
| wild-type | GCT/GTAAGT | wild-type | GCT/GTAAGT |
| C1 | ---/C----- | G4A5 | ---/---GA- |
| A2 | ---/-A---- | CAA | ---/---CAA |
| C3 | ---/--C--- | CTC | ---/---CTC |
| C4 | ---/---C-- | TCTC | ---/--TCTC |
| G4 | ---/---G-- | | |
| A5 | ---/----A- | | |
| A6 | ---/-----A | consensus | MAG/GTRAGT |

FIG. 1. Alternative splicing of hGH.FSX reporter pre-mRNA and mutations in the 5′ss signal. The generic expression vector phGH.FSX was previously described (5). It generates a chimeric hGH pre-mRNA containing an additional exon (0) and intron at the 5′ end, the alternative H-*ras* exon IDX and flanking sequences interrupting the intron between exons 3 and 4, and a heterologous poly(A) region. Untranslated exon regions (white boxes), hGH exons (gray boxes), and IDX with a previously described single-base deletion to permit translation readthrough (4) (black box) are shown. IDX skipping and inclusion are represented schematically below the hGH.FSX diagram and the mRNAs are named after the encoded proteins, 22kD and 22kD+, respectively. The DNA sequences of the 5′ss signal of intron C2 and the mutations mentioned in this paper are also shown, with slashes marking the cleavage position and dashes indicating bases identical to those in the wt sequence. The 5′ss consensus sequence is included for comparison (M, C or A; R, A or G).

U1 may contribute by selecting the local environment (17). Evidence that base pairing with U1 is not the critical reference for defining the 5′ss position first emerged from experiments with yeast cells where a mutation in the 5′ss signal resulted in the use of an aberrant cleavage site. Unexpectedly, when a U1 snRNA with a complementary change to restore base pairing at the authentic position was coexpressed, cleavage at the aberrant site was stimulated, not eliminated, despite the absence of significant complementarity between the altered U1 RNA and the sequence at the aberrant site (34–36). Our own work has extended these observations to intact mammalian cells and a more complex substrate (7; see below). In addition, two recent publications have shown in vitro that SR proteins in excess can rescue splicing at the normal position and promote aberrant splicing when the 5′ end of U1 RNA is unavailable for pairing with the pre-mRNA substrate (9, 42).

We are using a multi-intron reporter gene, termed hGH .FSX (Fig. 1), to study the contributions of base pair interactions to splice site selection in vivo. The reporter gene includes the coding exons for human growth hormone (hGH) and contains an exon (IDX) and flanking intron sequences from the human H-*ras* proto-oncogene within the intron separating hGH exons 3 and 4. IDX is normally included in the spliced transcript, which results in the production of an hGH-*ras*-hGH fusion protein, termed 22kD+ (7). However, mutations in the 5′ss signal at the end of IDX, such as G1 to C, A4 to G, and G5 to A (C1, G4, and A5 in Fig. 1), cause IDX skipping and result in the production of a smaller protein, referred to here as 22kD, which is authentic hGH (7); exon skipping due to mutations in the 5′ss signal is an indication of exon definition, i.e., the coordinate selection of exon boundaries (30). We previously showed that the G4 and A5 mutations could be suppressed and IDX inclusion could be restored by coexpression of U1 snRNAs, termed shift-U1s, with complementarity to sequences other than the mutated 5′ss signal, in particular any sequences tested within the adjacent exon or intron (7). Like the earlier observations in yeast cells (34–36), these results suggested that pairing of U1 with the 5′ss signal may not be necessary for initial recognition of the 5′ss region or for the catalytic steps of the reaction, i.e., 5′ss use. Given the presence of endogenous U1 RNA in our transfected cells, this conclusion was tentative, but the results clearly pointed to a function for U1 that requires base pairing with pre-mRNA, but not necessarily with the 5′ss signal or any other sequence in particular (7).

We now present results indicating that base pairing of U1 RNA with the 5′ss signal or sequences nearby stimulates splicing of the upstream intron, at the other side of the exon, whether or not the downstream intron is spliced. Thus, the normal 5′ss signal functions as a splicing enhancer for the upstream intron; this requires base pairing of the signal with U1, and other sequences in the area can assume this function by pairing with U1. In contrast, base pairing between the 5′ss signal and U1 appears to play a minor role at best in splicing of the downstream intron and may not be necessary at all for 5′ss recognition. We relate these findings to the possible role of U1 in exon definition.

## MATERIALS AND METHODS

**Plasmids.** The generic expression vectors phGH.FSX, pU1, and pDNase were described previously along with the approaches used to introduce mutations in the 5′ss signal of intron C2 in the hGH.FSX gene or toward the 5′ end of the U1 sequence in pU1 (5, 7). Details for specific mutations will be provided on request. The pGT+ and U6 snRNA vectors of Fig. 7 are described elsewhere (17).

phGH.FSX:A5:CGA (see Fig. 2B) was constructed as follows. A 160-bp fragment extending from the *Nhe*I site in intron C1 to the first *Sma*I site of intron C2 was obtained from phGH.FSX:A5 (5). A second fragment was generated by PCR with a forward primer [CGA(15-17)] initiating with the first nucleotide after the first *Sma*I cleavage site of intron C2 and specifying the ACG→CGA substitution at positions 15 to 17 of the intron and a reverse primer (Ras/hGH2860) that overlaps the *Bst*EII site downstream in the intron; phGH.FSX:A5 (5) was used as the template. The PCR material was treated with Klenow DNA polymerase and digested with *Bst*EII. The two fragments were ligated to a vector fragment obtained by digestion of phGH (5) with *Nhe*I and *Bst*EII.

phGH.FSX:A5:Yn.up (see Fig. 6A) was created by insertion of two PCR-generated fragments between the *Nhe*I and *Bst*EII sites in the intron C linker of phGH (5). With phGH.FSX:A5 as a template, the first fragment extended from within exon 3 to the polypyrimidine tract of intron C1 [primers hGH1718 and Yn:opt(−)]; the small piece released by digestion at the *Nhe*I site in the intron was isolated. The second PCR fragment generated with the same template extended from the polypyrimidine tract of intron C1 to the *Bst*EII site of intron C2 [primers Yn:opt-2(+) and Ras/hGH2860]; this fragment was digested with *Bst*EII. Together, the two Yn:opt primers specified the desired changes. pGT+: Yn.up was constructed by replacing the region of pGT+ between *Nhe*I and *Bgl*I sites in intron C1 and IDX, respectively, by a corresponding fragment from phGH.FSX:A5:Yn.up.

p22k, p22k+, and pcryp+12 (previously called p22k.cryp) have been described elsewhere (7). Cytoplasmic RNA from 293 cells cotransfected with phGH.FSX: CTC · GA (17) and pU1-α+14 was used for the construction of pcryp2+12 and pcryp2+22 by the same procedure as previously described for pcryp+12 (7); phGH.FSX:CTC · GA has five mutations at the start of intron C2, which prevents splicing at the normal position in the presence of U1-α+14 (17). Thus, pcryp2+12 differs from pcryp+12 by the CTC · GA mutation and G instead of C at the first intron C2 position. pcryp2+22 is identical to pcryp2+12, except that it has the first 22 nt of intron C2 instead of the first 12. p22k+C2 (see Fig. 4B) has the CTC · GA version of intron C2 inserted between IDX and exon 4 of p22k+. It was constructed by replacing the IDX-exon 4 region of p22k+ between *Bgl*I and *Bst*XI sites by the corresponding intron C2-containing region of phGH.FSX:CTC · GA.

pDNaseΔ10 has a 10-bp deletion in the 5′ untranslated region of pDNase between the transcription start site and the binding site for the RKΔI oligonucleotide used in the primer extension assay (see below); details are available on request. pDNase ^ E5 was constructed by transferring the *Bgl*II-*Pvu*II fragment of hGH exon 5, upstream of the stop codon, from phGH into the *Hin*dIII site in the 3′ untranslated region of pDNaseΔ10; the orientation of the insert relative to the direction of transcription is the same as that in phGH.

The integrity of new PCR-derived regions in our constructs, including desired mutations, was routinely confirmed by DNA sequencing through the entire region(s).

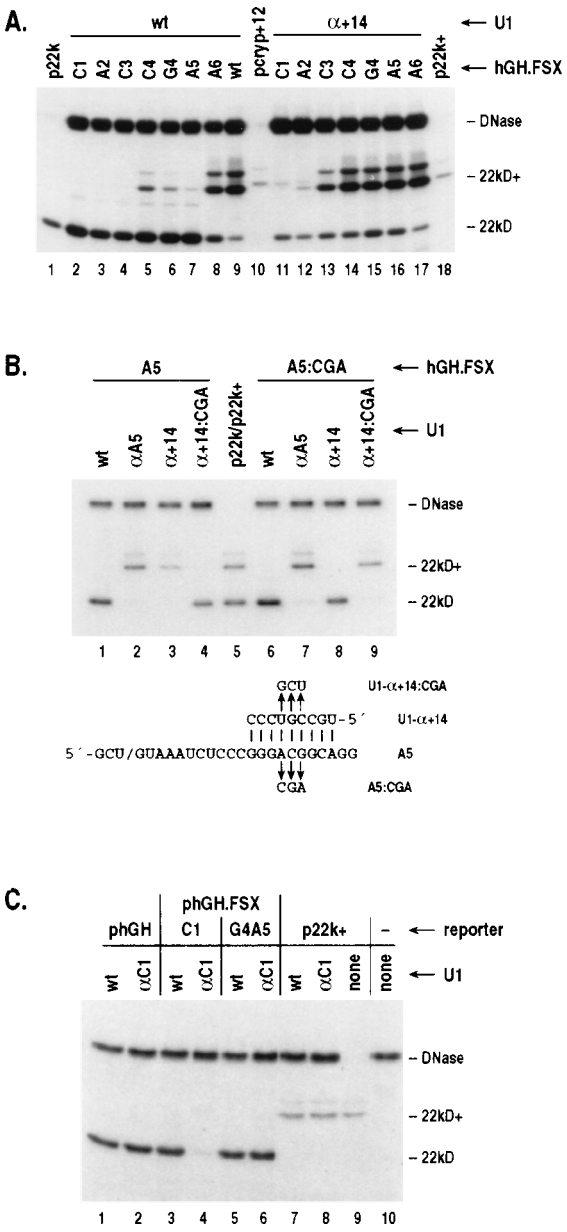



FIG. 2. Protein assays to distinguish IDX inclusion and skipping in transiently transfected 293 cells. Cells were incubated 2 days after transfection with a mixture of $^{35}$S-labeled methionine and cysteine, the media were immunoprecipitated with antisera specific for hGH and DNase proteins, and the precipitates were analyzed on sodium dodecyl sulfate-acrylamide gels, as previously described (5). (A) hGH.FSX expression vectors with the single-base changes in the 5′ss signal of intron C2 indicated above the lanes were cotransfected with the previously described wt or mutant U1 plasmids pU1-wt and pU1-α+14 together with the reference vector pDNase (5, 7). Control transfections were performed with vectors expressing cDNAs corresponding to 22kD mRNA (p22k), 22kD+ mRNA (p22k+), and cryp+12 mRNA (pcryp+12; previously called p22k.cryp) (7). (B) To demonstrate that the suppression of 5′ss mutations by U1-α+14 requires base pairing with its complementary sequence in intron C2, the central 3 nt of the intron C2 sequence were changed in the A5 reporter (ACG→CGA) to create the variant termed A5:CGA and the complementary changes (CGT→TCG) were introduced into pU1-α+14 to generate the U1-α+14:CGA gene, as illustrated at the bottom of the panel. The A5 sequence shown corresponds to the last three bases of IDX followed by a slash and the sequence of intron C2, including the G-to-A change at position 5. The complementary sequence at the 5′ end of U1-α+14 is shown in the reverse orientation above the A5 sequence, and complementary bases are indicated by vertical lines. The arrows point to the mutations in the U1-α+14 gene and the A5 reporter gene, respectively. At the top of the panel, the U1 RNAs coexpressed with each reporter construct and pDNase included the previously described U1-αA5, which is complementary at its 5′ end to the first nine bases of the A5 sequence shown (7). A cotransfection of p22k and p22k+ was included for reference. (C) Specificity of U1-αC1-mediated reduction in reporter protein production. phGH contains a short linker in the hGH intron between exons 3 and 4 instead of IDX and flanking sequences derived from the H-*ras* gene (7). Control transfections were performed with p22k+ alone (lane 9) and pDNase alone (lane 10).

**Transfections and protein assay.** 293 and HeLa cells were transfected with 5 to 10 μg of phGH.FSX, 4 to 8 μg of pDNase, pDNaseΔ10, or pDNase^E5, and 10 to 20 μg of pU1 by the calcium phosphate procedure, essentially as described previously (6). The experiment of Fig. 7 was performed under conditions similar to those detailed in reference 17. Metabolic labeling, immunoprecipitation, and subsequent manipulations were performed as previously described (5).

**RNA analysis.** Cytoplasmic RNA isolation and primer extension were performed as described previously (40). The RKΔI primer (see below) recognizes sequences in the 5′ untranslated regions of hGH.FSX and DNaseΔ10 mRNAs that differ by 10 nt in their positions relative to the transcription initiation site.

The procedure and conditions for S1 nuclease protection were as previously described (4); annealing was at 50°C. Whenever a probe included a site altered by mutation in the transfected DNA, the probe was made to have the same mutation(s).

The internally labeled probes of Fig. 4A were synthesized in the presence of [α-$^{32}$P]dATP on phagemid-derived single-stranded templates with primer Ras/hGH2860, which overlaps the *Bst*EII site at the junction of *ras* and hGH sequences in intron C2 (see "Plasmids" above and see below; also references 4 and 5). Each phagemid contained a 331-bp *Nhe*I-*Pst*I fragment, initiating within intron C1 and extending to the end of intron C2, from hGH.FSX genes with the different 5′ss mutations inserted between the *Xba*I and *Pst*I sites of pUC118 (46). The synthesis products were digested at the unique *Eco*RI site in the pUC linker, and the labeled strand was isolated as described previously (4).

To generate probes 1 of Fig. 5B and 7, PCR fragments from the beginning of IDX to the Ras/hGH2860 primer site in intron C2 were first cloned into the *Sma*I site of pUC119; in the process, mutations were introduced to create a *Bss*HII site at the start of the IDX sequence (details available on request). Recombinants that had the unique *Hin*dIII site of pUC119 near the intron C2 end of the insert were used to isolate double-stranded probes 3′ end labeled with [α-$^{32}$P]dCTP at the *Bss*HII site with a pUC119-derived extension to the *Hin*dIII site at the other end. The other 3′-end-labeled probes used in this work were generated in a similar manner. In Fig. 4B, a PCR fragment extending from upstream of the *Bst*EII site in intron C2 to the end of the binding site of primer hGH2444 some 115 nt into the next intron was inserted into the *Sma*I site of pUC119, and a recombinant in the proper orientation was used to prepare the double-stranded probe, which was labeled at the *Bst*EII site and had a pUC-derived extension to the *Hin*dIII site at the other end. For the exon 5-specific probe (probe 2 in Fig. 5A to C and Fig. 7), a 176-bp *Bgl*II-*Eco*RI fragment containing a portion of exon 5 from the hGH.FSX gene was cloned between the corresponding sites of pSP72 (Promega Corp., Madison, Wis.). The probe was labeled at the *Bgl*II site and extended to the *Hin*dIII site in the pSP72 sequence at the other end of the insert.

A PCR strategy was used to generate the 5′-end-labeled, double-stranded probes of Fig. 5A and C. Appropriate template plasmids were used to direct the addition of unrelated sequences to the 3′ ends of the probes. In Fig. 5A, the template was one of the phagemid constructs described for the probes of Fig. 4A above. A region initiating in the upstream flanking pUC sequence and extending through the downstream portion of intron C1 to the *Sma*I site in IDX was amplified with pUC-specific forward primer 118-1004 and an IDX-specific reverse primer, Ras2676(−). The PCR product was 5′ end labeled with polynucleotide kinase and [γ-$^{32}$P]ATP and digested at the *Hin*dIII site in the pUC linker, and the downstream fragment was gel purified. In Fig. 5C, the template plasmid contained a 670-bp *Bst*EII-*Eco*RI fragment from phGH.FSX, extending from intron C2 to just past the stop codon of exon 5, between the *Sma*I and *Eco*RI sites of pUC119. A 5′ vector-insert region was amplified with forward primer 118-1004 and exon 4-specific reverse primer hGH1513, the product was 5′ end labeled and digested with *Hin*dIII, and the downstream fragment was isolated.

**Oligonucleotide sequences.** The sequences of oligonucleotides specifically mentioned here are given below (5′→3′); other sequences are available on request. The oligonucleotide sequences are as follows: CGA(15-17), GGGCGAG CAGGGCAGTGAG (intron C2) (forward); Ras/hGH2860, CCGAGGTCACC GCTCCGGCCTGGCTCAG (intron C2) (reverse); hGH1718, TTCTCAGAGT CTATTCCGACACC (exon 3) (forward); Yn:opt(−), GAAGAGAAAGAGT CAGTGAGTGCTG (intron C1) (reverse); Yn:opt-2(+), CTTTTTTTTTTCAC AGGGCAGCCGCTC (intron C1-IDX junction) (reverse); RKΔI, ACACGT CACTCTTGGCACG (exon 0) (reverse); hGH2444, GTGAGTTCTCTTGGG TCAG (intron D) (reverse); 118-1004, CACAGGAAACAGCTATGACCATG (pUC); Ras2676(−), GGGGGTCCCAGAGGGTCC (IDX) (reverse); and hGH1513, AGAGGCGCCGTACACCAGGCTGT (exon 4) (reverse).

## RESULTS

**Suppression of individual point mutations.** Figure 2A illustrates the suppression of single-base changes in the IDX-associated 5′ss signal by a typical shift-U1 RNA, U1-α+14, which is complementary near its 5′ end to positions 12 to 20 of intron C2 of the hGH.FSX reporter (7). The results were obtained by immunoprecipitation of media from cells pulse-labeled 2 days after transfection with antisera specific for hGH and a reference protein, DNase, expressed from a cotransfected control vector (5). The 22kD and 22kD+ proteins indicate IDX skipping and accurate IDX inclusion, respectively, and cDNAs expressing each protein individually were included for comparison (Fig. 2A, lanes 1 and 18, respectively); the use of an aberrant 5′ss 12 bases downstream of IDX results in a protein termed cryp+12 (22kD-cryp in reference 7), and the corresponding cDNA was expressed in lane 10. Each of the hGH-related proteins is produced as two distinct forms. The larger form probably corresponds to uncleaved precursor, and the smaller, more abundant form is marked at the side of this and other figures.

The normal hGH.FSX vector produced mainly the 22kD+ protein encoded by IDX-containing mRNA (Fig. 2A, lane 9), but this protein was no longer produced when positions 1 to 3 were changed individually (lanes 2 to 4; mutations shown in Fig. 1) and more 22kD protein, signifying IDX skipping, was generated instead. Thus, the mutations prevented use of both the 5′ss at the end of IDX and the 3′ss at the end of the preceding intron, suggesting that normal IDX splicing involves an exon definition step (30). For individual mutations at positions 4 to 6, IDX inclusion was reduced to different degrees and skipping was increased proportionally (lanes 5 to 8). Coexpression of U1-α+14 suppressed the individual mutations at positions 3 to 6 rather efficiently (lanes 13 to 17), although small differences are apparent. In each case, the 22kD+ signal was increased and the 22kD signal was reduced compared with the corresponding U1-wt lane, suggesting that base pairing with U1 at a distance from the 5′ss signal restores splicing of both the downstream and upstream introns. However, the opposite effects of U1-α+14 on splicing and skipping appeared to be uncoupled in the C1 and A2 situations. While U1-α+14 reduced the abundance of the IDX-skipping product (22kD) here as well, no (C1) or little (A2) 22kD+ protein was produced (lanes 11 and 12) and only a small amount of aberrant cryp+12 product, migrating just above the 22kD+ protein, was detectable (compare with lane 10). Thus, U1-α+14 reduced the overall reporter signal. We have previously described a similar observation for the C1 mutant coexpressed with a U1 RNA matching the mutated 5′ss sequence itself at all nine positions (U1-αC1; reference 7 and see below). The simplest explanation of these results is that both U1 RNAs with complementarity to the 5′ss sequence and shift-U1 RNAs establish a commitment to IDX inclusion, thereby preventing IDX skipping, but that a protein product recognizable by our antiserum is generated only when an acceptable 5′ss signal is available.

The control experiment in Fig. 2B shows that the effect of U1-α+14 depends on base pairing of the U1 RNA with its complementary sequence in intron C2. Thus, when positions 15 to 17 of intron C2 in the A5 vector were changed (ACG to CGA; Fig. 2B, bottom), U1-α+14 no longer promoted synthesis of the 22kD+ protein (compare lanes 3 and 8), but this defect was suppressed by compensatory changes in U1-α+14 (lane 9; the compensatory changes are shown below the gel). U1-αA5, which is complementary to the mutated 5′ss signal itself, was included as a control, and it can be seen that the ability of this U1 RNA to promote 22kD+ synthesis was not affected by the mutations at positions 15 to 17 of the reporter gene (lanes 2 and 7). Similar results have been obtained for all other shift-U1 RNAs tested in a comparable manner, and an example has been presented before (7).

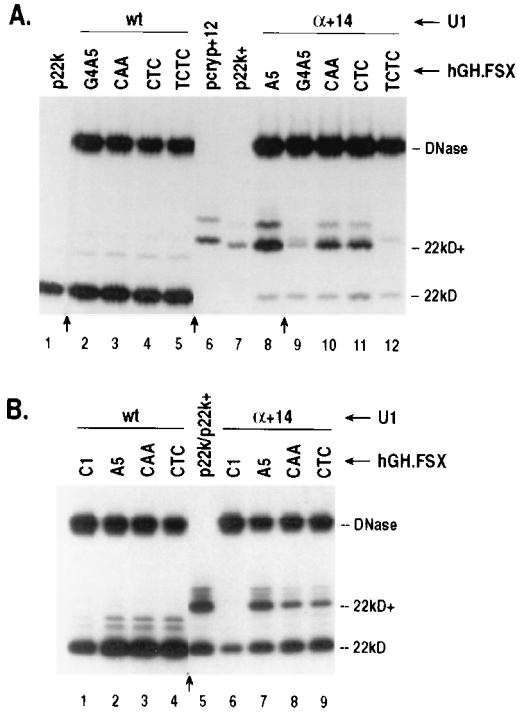



FIG. 3. Examination by protein assay of reporter hGH.FSX substrates with multiple-base changes in the IDX-associated 5′ss signal for U1-α+14-promoted IDX inclusion in transfected 293 (A) and HeLa (B) cells. As before, reporter cDNA expression vectors were transfected as controls either alone (lanes 1, 6, and 7 in panel A) or in combination (lane 5 in panel B). Vertical arrows at the bottom of the gels indicate where irrelevant lanes were deleted.

In Fig. 2C, U1-αC1 was used to demonstrate that the loss of reporter signal observed in some cases is a specific effect. It can be seen that U1-αC1, similar to U1-α+14 in Fig. 2A, dramatically reduced the amount of 22kD protein produced by the C1 gene and no alternative protein appeared instead (lanes 3 and 4). In contrast, no loss of reporter signal was evident when the 5′ss sequence at the end of IDX was missing (compare lanes 1 and 2 and lanes 7 and 8; see legend for details) or differed from the C1 sequence at three positions (lanes 5 and 6). Together, these control experiments (Fig. 2B and C) demonstrate that the observed effects depend on base pairing between the transfected U1 RNAs and their intended target sequences on the reporter pre-mRNA.

**Suppression of combined point mutations.** On the basis of the results of Fig. 2A, reporter genes with multiple-base changes at positions 3 to 6 were constructed and tested for accurate splicing in the presence of U1-α+14 (Fig. 3A). The triple mutations at positions 4 to 6 (AGT to CAA or CTC; Fig. 1) reduced complementarity to endogenous U1 to three of nine positions, and the altered sequences lack significant complementarity to U1-α+14. Nonetheless, the 22kD+ protein was produced in the presence of U1-α+14 (compare lanes 10 and 11 with lanes 3 and 4), indicating that both introns C1 and C2 were accurately spliced. To determine whether this might be specific for 293 cells, which are known to express abnormally high levels of the SR protein SF2/ASF (12), a similar experiment was performed with HeLa cells. The results

were qualitatively the same (Fig. 3B, lanes 3 and 4 and lanes 8 and 9), although the effect of U1-α+14 was less dramatic overall. These data show that substantial base pairing between the 5′ss sequence and U1 snRNA, i.e., conventional base pairing at more than three positions, is not essential for recognition or use of the correct 5′ss at the end of IDX. They also indicate that accurate splicing in this situation is not dependent on increased levels of SF2/ASF.

The two other new mutants analyzed in Fig. 3A showed more-severe defects (lanes 9 and 12). This is surprising for the G4A5 mutation (lane 9), given the mild effects of the G4 and A5 mutations individually (Fig. 2A). The G4A5 and TCTC mutations are not suppressed by U1 RNAs with complete complementarity to the altered 5′ss sequence in each case (unpublished data), which indicates these mutations are not merely more severe than the G4 and A5 or CTC mutation, respectively, because of the additional mismatch in each case with endogenous U1. For the TCTC mutation, it is possible that the T at position 3 is largely responsible for the severe defect, but further studies are required to clarify the situation in both cases.

**Base pairing of U1 with the downstream intron promotes splicing of the upstream intron.** To better characterize the effect of U1-α+14 on IDX splicing, S1 nuclease protection analyses were performed with the internally labeled probe shown schematically at the top of Fig. 4A. For reference, RNA samples were also compared by primer extension (Fig. 4A, bottom) with a primer that recognizes both the first exon of the reporter (exon 0) and a sequence near the 5′ end of the DNase control RNA.

The S1 analysis (Fig. 4A, center) shows that a common new product, labeled IR, was formed when genes with different mutations in the 5′ss signal were expressed together with U1-α+14 (compare lanes 7, 8, 11, and 12 with lanes 2 to 5). Control experiments demonstrated that the IR product contained IDX and the intron C2 portion of the probe but lacked intron C1 (data not shown), suggesting that the product was due to splicing of intron C1 only; IR stands for intron C2 retention. The experiment in Fig. 4B was performed to determine whether the responsible RNA was spliced at the 3′ end of exon 4. The probe, illustrated at the bottom of Fig. 4B, had an end label in intron C2 and extended into the next intron downstream; a 22kD+ cDNA expression vector that contains but cannot splice intron C2 was analyzed for comparison (lane 3). The results show that coexpression of the C1 reporter and U1-α+14 (lane 2) resulted in the appearance of a protected product of the same size as the IR product obtained with the splicing-defective cDNA (lane 3). It can be concluded that the IR products in Fig. 4A and B result from protection by an RNA that retains intron C2, but not intron C1 or D, and hence, that U1-α+14 promotes the inclusion into spliced RNA of a composite exon consisting of IDX, intron C2, and exon 4. The size of the composite exon is 427 nt, which is significantly more than the preferred size of internal exons (<300 nt; see reference 2 for a review).

In the S1 analysis of Fig. 4A, lane 10 shows the position of the fragment protected by fully spliced 22kD+ mRNA, i.e., IDX (smallest product marked to the side of the gel); the RNA was from cells transfected with a normal 22kD+ cDNA vector. As expected, products of the same size were observed in the lanes for the normal (wild-type [wt]) reporter gene coexpressed with U1-wt (lane 1) or U1-α+14 (lane 6), and little or no IR product was detected. Thus, IDX-containing mRNA was fully spliced. RNA produced by the A5 gene in the presence of U1-α+14 also protected IDX, but some IR product was formed as well (lane 7). In contrast, IR product was the main

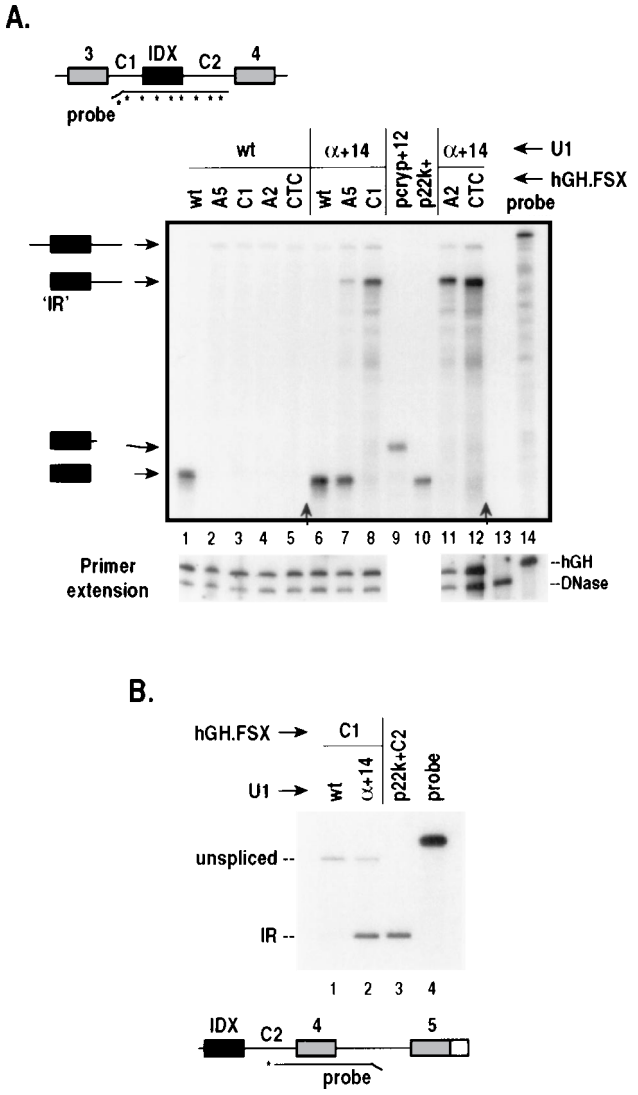



FIG. 4. S1 nuclease protection analyses of cytoplasmic RNAs from transiently transfected 293 cells. In both panels, the region of the hGH.FSX gene present in the probe is indicated by a horizontal line below a representation of the relevant portion of the gene; the presence of an unrelated sequence at one end of the probe is indicated by a diagonal extension. Internal or end label are symbolized by multiple asterisks underneath or a single asterisk at one end, respectively. (A) The main products of the S1 analysis at the center are identified at the left, with the black box representing IDX and the horizontal lines representing flanking introns. 'IR', intron C2 retention product (see text). Mutant-specific probes were used to avoid cleavage due to mutations in the 5′ss signal; 1/30 of the amount of probe used for the A2 RNAs was loaded in lane 14. As before, positions where irrelevant lanes were deleted are marked by vertical arrows at the bottom of the gel. At the bottom, reporter and DNase RNAs were detected by primer extension as products of 126 and 116 nt in length, respectively. The single 5′-$^{32}$P-labeled primer was complementary to a sequence in the first exon of the hGH.FSX genes (exon 0 [Fig. 1]); the same recognition sequence is present in the cotransfected pDNaseΔ10 plasmid at a position 10 nt closer to the transcription initiation site (see Materials and Methods for details). The last two lanes, which do not correspond to lanes on the S1 gel, contain RNAs from cells transfected with pDNaseΔ10 alone and phGH.FSX:CTC alone. (B) Protection of the entire hGH.FSX portion of the probe indicates retention of the two introns flanking exon 4 and suggests the presence of unspliced RNA. IR, selective retention of intron C2, p22k+C2 is p22k+ with intron C2 regenerated between IDX and exon 4; splicing is prevented by five base changes at the start of the intron (see Materials and Methods). A fraction (1/30) of the amount of probe used in the S1 reactions was electrophoresed in the lane labeled probe.

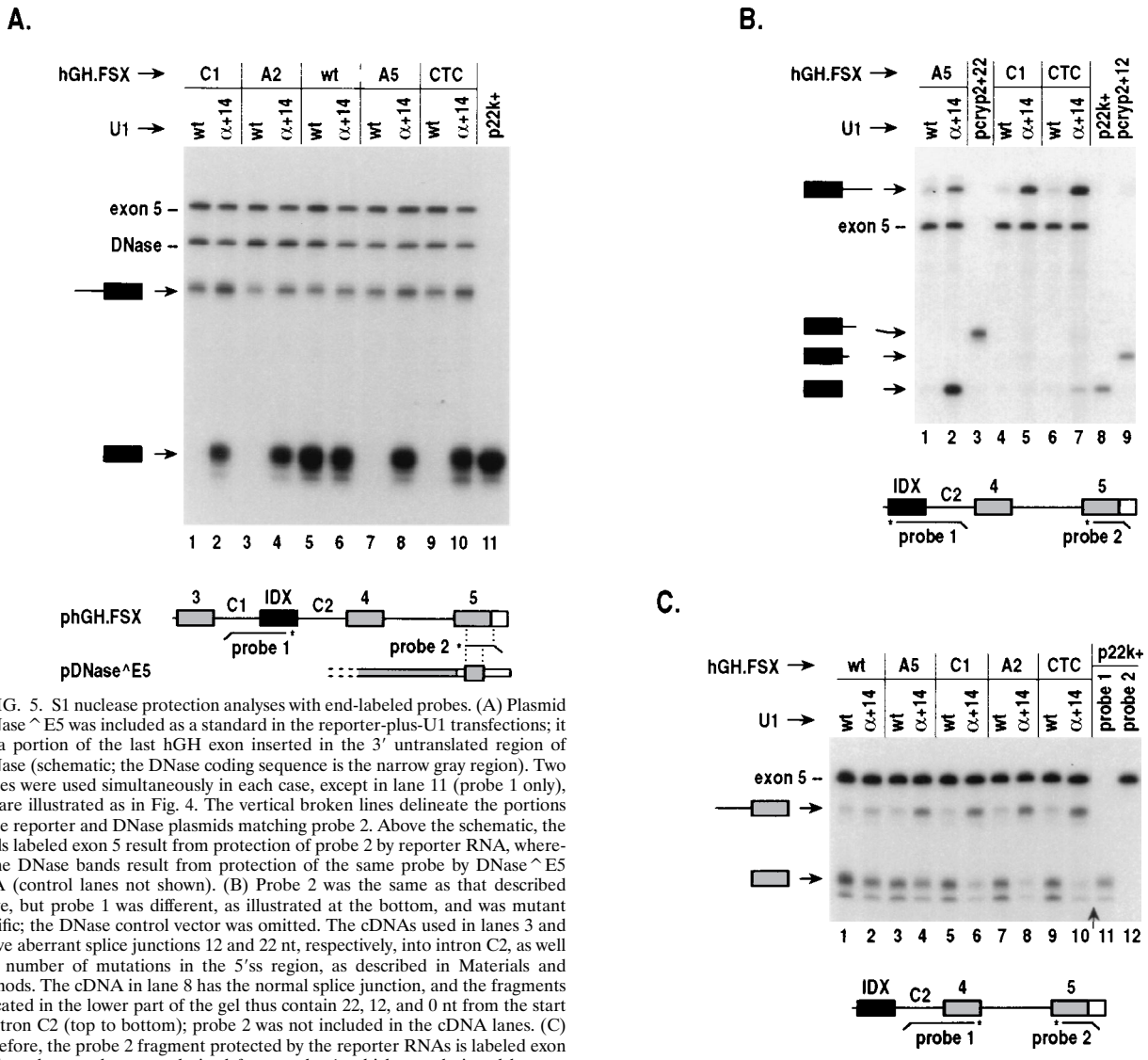



FIG. 5. S1 nuclease protection analyses with end-labeled probes. (A) Plasmid pDNase ^ E5 was included as a standard in the reporter-plus-U1 transfections; it has a portion of the last hGH exon inserted in the 3′ untranslated region of pDNase (schematic; the DNase coding sequence is the narrow gray region). Two probes were used simultaneously in each case, except in lane 11 (probe 1 only), and are illustrated as in Fig. 4. The vertical broken lines delineate the portions of the reporter and DNase plasmids matching probe 2. Above the schematic, the bands labeled exon 5 result from protection of probe 2 by reporter RNA, whereas the DNase bands result from protection of the same probe by DNase ^ E5 RNA (control lanes not shown). (B) Probe 2 was the same as that described above, but probe 1 was different, as illustrated at the bottom, and was mutant specific; the DNase control vector was omitted. The cDNAs used in lanes 3 and 9 have aberrant splice junctions 12 and 22 nt, respectively, into intron C2, as well as a number of mutations in the 5′ss region, as described in Materials and Methods. The cDNA in lane 8 has the normal splice junction, and the fragments indicated in the lower part of the gel thus contain 22, 12, and 0 nt from the start of intron C2 (top to bottom); probe 2 was not included in the cDNA lanes. (C) As before, the probe 2 fragment protected by the reporter RNAs is labeled exon 5. The other products are derived from probe 1, which was designed here to detect splicing at the end of intron C2 (see schematic). Two of these products are marked at the side of the gel by illustrations representing exon 4 alone or together with a portion of intron C2; the third (smallest band, not labeled) results from aberrant cleavage of probe 1 (compare with lane 11; also data not shown). Protection of the two probes separately by RNA from cells transfected with p22k+ was analyzed in the last two lanes at the right; the vertical arrow indicates deletion of irrelevant lanes.

S1-resistant fragment for the genes with more-severe mutations in the 5′ss signal in the presence of U1-α+14 (lanes 8, 11, and 12). These results show that U1-α+14 consistently restored use of the upstream 3′ss, whether or not splicing occurred at the downstream 5′ss.

The results suggest a simple explanation for the overall loss of protein signal noted earlier for some of the mutant genes in the presence of U1-α+14. Thus, where U1-α+14 causes retention of intron C2, in-frame termination codons are introduced between IDX and exon 4, resulting in a translation product that may be unstable or not recognized by our hGH-specific antiserum. Other explanations are less attractive. For example, it is not likely that the loss of signal was caused by decreased RNA stability or a decrease in the rate of RNA export from the nucleus when intron C2 is retained, because this would decrease the total amount of cytoplasmic reporter RNA, which is not observed (Fig. 4A, bottom). Similarly, a loss of signal due to abortive splicing, i.e., a block in the second step of the reaction, would show as a reduction in the amount of cytoplasmic reporter RNA, since dead-end splicing intermediates are not exported from the nucleus.

To strengthen the conclusions from Fig. 4, end-labeled probes were used to measure splicing at the two ends of IDX and the downstream 3′ss separately. In Fig. 5A, the U1 and hGH.FSX genes were cotransfected with a cDNA control vector expressing a chimeric DNase transcript that contains a portion of hGH exon 5 (pDNase ^ E5). Total cytoplasmic reporter and control RNAs were visualized by an hGH exon 5-specific probe (probe 2; the products are labeled exon 5 and DNase, respectively), and use of the 3′ss of intron C1 was measured with a second end-labeled probe (probe 1), giving rise to the two other products labeled at the side of the gel. Lane 11 serves as a marker for use of the 3′ss of intron C1; the RNA was from cells expressing 22kD+ cDNA, and probe 2 was omitted. The band of lane 11 is also seen in lane 5 for the wt gene, but not in lanes 1, 3, 7, and 9 for the mutated genes, which confirms that mutations in the 5′ss signal downstream

abolish use of the 3′ss of intron C1. In the presence of U1-α+14, however, use of the 3′ss was restored (lanes 2, 4, 8, and 10). These results indicate that use of the 3′ss is normally supported by base pairing between endogenous U1 and the 5′ss signal downstream, that mutations in the downstream signal, including single-base changes, prevent base pairing with endogenous U1, thereby eliminating use of the upstream 3′ss, and that base pairing of U1 with a sequence in intron C2 can compensate for the lost interaction with the 5′ss signal. Together with the earlier protein data, these results strongly suggest that the normal 5′ss signal acts as an enhancer of intron C1 splicing, that this activity is due to pairing of the signal with endogenous U1, and that pairing of U1 with a different sequence in the area is sufficient to reproduce the activity.

Figure 5A also strengthens the suggestion from the primer extension data of Fig. 4A that U1-α+14 does not affect total cytoplasmic reporter RNA levels, since the ratios of two largest bands are not significantly altered by U1-α+14. This indicates that the exon 5 signal to measure total reporter RNA levels provides an adequate standard for normalization in comparisons of U1-α+14 and U1-wt, which is noted here because the DNase control vector was not always included in the experiments presented below.

In Fig. 5B, a 3′-end-labeled probe spanning the IDX-intron C2 junction was used to measure splicing at the end of IDX (probe 1 in the schematic); probe 2 was used to detect hGH exon 5, but no DNase vector was included. The largest band marked at the side results from protection of IDX and intron C2 and probably represents unspliced RNA in the U1-wt lanes (lanes 1, 4, and 6). The band is more intense in the presence of U1-α+14 in each case (lanes 2, 5, and 7), and the results of Fig. 4 indicate that this is caused by inclusion of IDX and intron C2 in spliced RNA. Protection of probe 1 to the 3′ end of IDX gave rise to the product in lane 8 (smallest band marked at the side of the gel), where RNA was used from cells transfected with the 22kD+ cDNA vector. The same product is seen where the A5 or CTC gene was coexpressed with U1-α+14 (lanes 2 and 7), but the intensity of the band is very different in the two situations and the product was not detected with the C1 reporter (lane 5). Thus, whereas U1-α+14 restores use of the upstream 3′ss irrespective of which mutation is present in the 5′ss signal (Fig. 5A), the mutations differ widely in their effects on 5′ss use.

Lanes 3 and 9 of Fig. 5B are controls showing protection of IDX up to aberrant 5′ splice sites 22 and 12 nt, respectively, into intron C2; the RNAs were from cells expressing cDNAs with the corresponding aberrant splice junctions. These lanes were included in the anticipation that mutations interfering with recognition or use of the normal 5′ss, such as C1 or CTC, would increase splicing at aberrant sites in the presence of U1-α+14. However, this is not observed (compare lanes 5 and 7 with lane 2), indicating that the 5′ss at the end of exon 4 is the preferred alternative when the 5′ss of intron C2 is defective. In other words, there is no significant tendency toward reducing the size of the composite exon incorporated into spliced RNA in these situations.

In Fig. 5C, use of the 3′ss of intron C2 in the presence or absence of U1-α+14 was compared. As before, hGH exon 5 was visualized for reference. The results demonstrate that U1-α+14 inhibits use of the 3′ss site of intron C2 (smallest band marked at the side of the gel, identified by comparison with lane 11) in favor of intron C2 retention (second band from the top). As expected, the effect is largest for the most-severe mutations (lanes 5 to 10), but a slight switch is apparent even for the wt reporter (compare lanes 1 and 2). Together with Fig. 5A, these results show that U1-α+14 can cause a dramatic

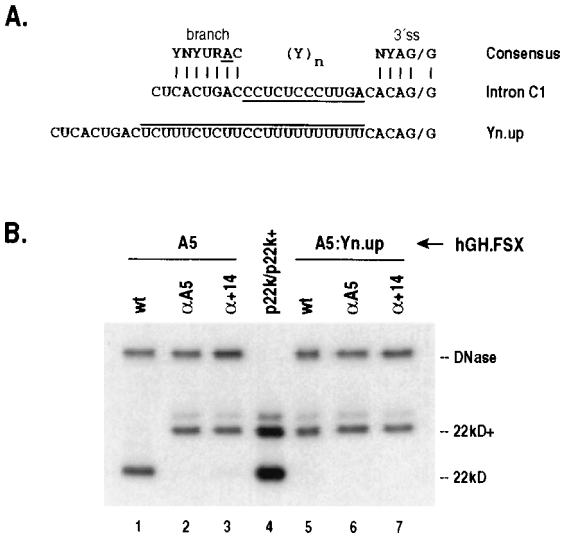



FIG. 6. Improvement of the 3′ss region of intron C1 and analysis of the effect by protein assay. (A) The sequence at the 3′ end of intron C1 is shown and bases matching those of the consensus sequences at the top for the branch point region (the branch point is underlined) and the 3′ss are indicated by vertical lines. The position of 3′ cleavage (/) is marked. $(Y)_n$ on the upper line represents the consensus sequence for the polypyrimidine tract (Y is C or U). Also at the top, N is any of the four ribonucleotides and R is A or G. At the bottom, the overlined stretch of pyrimidines was substituted for the underlined sequence of intron C1 to create the Yn.up mutation. (B) Comparison of the normal A5 vector and its counterpart carrying the Yn.up mutation in 293 cells. See the legend to Fig. 2B for a description of U1-αA5. pDNase was included in the transfections, except in the cotransfection of the two reporter cDNAs (lane 4).

change in 3′ss choice, with increased use of the upstream site at the end of intron C1 accompanied by decreased use of the downstream site at the end of intron C2. Thus, base pairing of U1 with a sequence in intron C2 promotes use of the 3′ss across the exon, but not the 3′ss across the intron.

**U1–pre-mRNA base pairing compensates for a weakness in upstream 3′ss signal but is not important for splicing of the downstream intron.** It has been shown previously that the incorporation of internal exons into spliced RNA is influenced by the strength of the polypyrimidine tract in the upstream intron (10) and that base pairing of U1 with the 5′ss sequence at the other end of the exon can compensate for weaknesses in the polypyrimidine tract (15). The longest uninterrupted stretch of pyrimidines in intron C1 is found in the 18-nt interval between the putative branch point and the 3′ cleavage site (Fig. 6A) and is separated by two purines and a pyrimidine from the consensus 3′ss signal (NYAG/G, where N is any nucleotide and Y is pyrimidine). To test whether this is a suboptimal arrangement necessitating a contribution of U1 for function, we changed the region in the A5 reporter between the putative branch site and the 3′ss signal of intron C1 to 24 pyrimidines (Fig. 6A, bottom). The protein assay results in Fig. 6B show that IDX inclusion, i.e., formation of the 22kD+ protein, was no longer dependent in this situation on coexpression of altered U1 RNAs (compare lanes 1 and 5). In addition, U1 RNAs that normally promote IDX inclusion, whether a shift-U1 (lane 3) or a U1 RNA with complete complementarity to the mutated 5′ss sequence of the A5 reporter (U1-αA5; lane 2), appeared to have no further stimulatory effect (lanes 7 and 6, respectively). We do not know whether the increased distance between the putative branch site and the 3′ss, the increased length of the pyrimidine tract, or both are necessary for the effect, but the results confirm that the sequence context of the

normal 3′ss is suboptimal. Furthermore, the results show that base pairing of U1 with sequences downstream can compensate for this weakness in the upstream signal.

The results also support the earlier suggestion that base pairing between the normal 5′ss sequence and U1 is necessary only for splicing of the upstream intron, but not for splicing of the downstream intron. To strengthen this suggestion, we used the CTC mutation to reduce the complementarity between endogenous U1 and the 5′ss sequence to three positions and asked whether significant splicing of intron C2 could be achieved in this case without restoring complementarity between U1 and the mutated 5′ss sequence. The vector used in this experiment, pGT+, had additional changes in intron C2 immediately downstream of the 5′ss sequence, which together with the CTC mutation, eliminate base pairing of the region with U6 snRNA; this is described elsewhere (17). The S1 protection results in Fig. 7B demonstrate that foreign U1 RNAs were not required for splicing of intron C2 when the polypyrimidine tract of intron C1 was improved (pGT+:Yn.up) and U6 could pair with the 5′ss region as a result of compensatory changes in a cotransfected U6 gene (U6-αCTC, [Fig. 7A]; Fig. 7B, lane 7); the dependence of intron C2 splicing on U6-αCTC (lanes 5, 7, and 9) is discussed in detail elsewhere (17). The altered U1 RNA used in this experiment, U1-αCTC (Fig. 7A), had complete complementarity to the mutated 5′ss sequence and was active, as indicated by the increased intensity of the largest product in lane 4 compared with lane 2. However, U1-αCTC could not alone restore splicing of intron C2, whether or not the sequence of the upstream intron was improved (lanes 4 and 8). In the presence of U6-αCTC, U1-αCTC was necessary for splicing of intron C2 when the upstream intron was unchanged (compare lanes 3 and 5), but not when it was improved (compare lanes 6 and 8, 7, and 9). These results agree with the previous conclusion that base pairing between the 5′ss sequence and U1 is normally necessary for splicing of the upstream intron, but they also show that splicing of the downstream intron can proceed when there is very little potential for base pairing between the two sequences. This strongly suggests that base pairing between the 5′ss sequence and U1 is not important for splicing of the downstream intron.

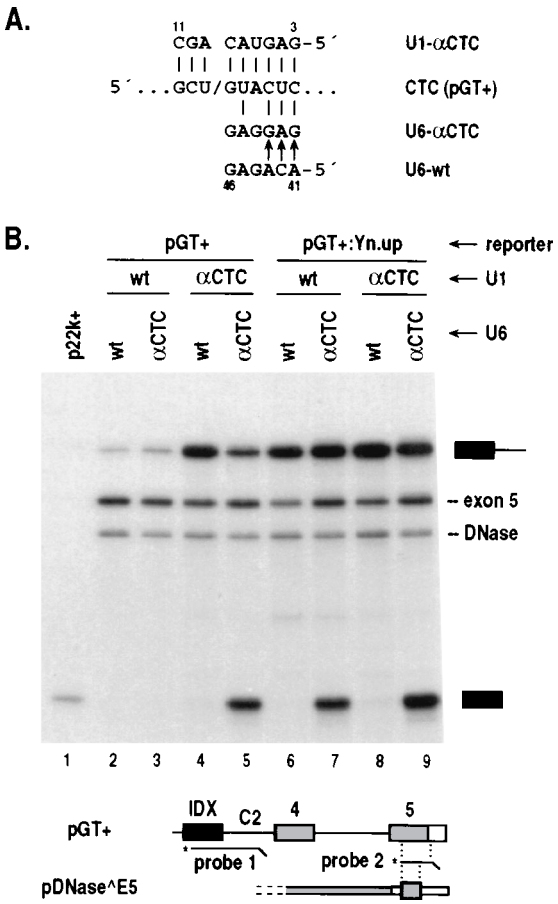



FIG. 7. Rescue of intron C2 splicing in the absence or presence of the Yn.up mutation by U1 and U6 RNAs with compensatory changes to the CTC mutation. The reporter vector, pGT+, differs from the previous CTC vector at positions 7 to 11 of intron C2, which are replaced by an unrelated 12-bp sequence (17; not shown). This eliminates complementarity between positions 8 and 9 of the intron and positions 39 and 38, respectively, of U6 snRNA, while the CTC mutation eliminates complementarity between positions 5 and 6 of the intron and positions 42 and 41 of U6 (17). (A) The 5′ss sequence (CTC) of pGT+ is shown 5′ to 3′, left to right. The lines underneath show the sequences at positions 41 to 46 of normal U6 RNA (U6-wt; 3′ to 5′) and U6-αCTC, which has mutations at positions 41 to 43 (vertical arrows) to permit base pairing with positions 4 to 6 of the CTC intron; complementary bases in the sequences are indicated by vertical lines. The top line shows the sequence at positions 3 to 11 of U1-αCTC (5′ to 3′, right to left), which is complementary to the reporter sequence and differs from U1-wt at positions 3 to 5 and 9 to 11. (B) S1 nuclease protection results with the two probes illustrated at the bottom. Normal or mutant U1 and normal or mutant U6 genes were cotransfected with the pGT+ reporter vector (lanes 2 to 5) or a pGT+ derivative with an improved polypyrimidine tract in intron C1 (pGT+:Yn.up; lanes 6 to 9); the pDNase ^ E5 vector was included for standardization (lanes 2 to 9). In lane 1, the RNA was from cells transfected with just the 22kD+ cDNA vector and probe 2 was omitted. The products are labeled as described above (e.g., legend to Fig. 5).

## DISCUSSION

The results in this paper demonstrate in vivo that base pairing of U1 with the 5′ss signal of an internal exon or a sequence in the downstream intron promotes removal of the upstream intron, irrespective of whether the downstream intron is retained or spliced. We show that the sequence at the upstream 3′ss is substandard and present evidence that U1 compensates for this weakness when it pairs downstream. Additional results raise the possibility that base pairing between U1 and the sequence at the 5′ss is not necessary for accurate splicing of the downstream intron.

**Base pairing with U1 downstream stimulates a weak upstream 3′ss.** In the absence of mutations, IDX inclusion is favored over skipping by our reporter gene, but the balance of inclusion versus skipping is completely reversed by different mutations in the IDX-associated 5′ss signal. We previously showed for the A5 mutation that this defect is suppressed by any of a number of altered U1 RNAs that can pair with fully complementary sequences within IDX, at the junction of IDX and intron C2, or within intron C2 (7); evidence was presented that this depends on base pairing between the complementary sequences. Our present work focuses on a single shift-U1 RNA, but no fundamental differences with other shift-U1 RNAs have been detected in related work presented elsewhere (17), indicating that our results here are not due to some unique property of U1-α+14 or the position of its complementary sequence.

We have shown that U1-α+14 increases splicing at the upstream 3′ss and believe this reflects bona fide stimulation of the upstream site, rather than some indirect effect. For example, it seems unlikely that the upstream site is used more efficiently because of inhibition by U1-α+14 of the downstream 3′ss at the end of intron C2, because use of the downstream site varies in the presence of U1-α+14, from complete

inhibition to almost normal activity (Fig. 5C), while the upstream site is always stimulated (Fig. 5A). It is possible, however, that U1 promotes stable pairing of the 5′ss and 3′ss of intron C2, whether or not the intron can be spliced, which could eliminate the option of IDX skipping and force splicing of intron C1. We do not favor this possibility, in part because there is a reasonable alternative to splicing of intron C1, which is skipping of the entire 427-nt exon composed of IDX, intron C2, and exon 4, and in part because the scenario relies on the untested idea that the 3′ss of intron C2 can be tied up in an unproductive interaction long enough and efficiently enough to achieve the necessary result. Instead, our findings that improvement of the polypyrimidine tract of the upstream intron has the same effect as U1-α+14 and that combining the two has no additional effect support the interpretation that U1-α+14 directly stimulates use of the upstream 3′ss. This is also suggested by the previous finding in vitro that pairing of U1 with a 5′ss signal at the end of an internal exon stimulates binding of $U2AF^{65}$ to a weak polypyrimidine tract in the upstream intron (15); this could be the responsible mechanism in our case. Thus, we favor the interpretation that base pairing of U1, whether with the 5′ss signal or with sequences in the adjoining exon or intron, compensates directly for a weakness in the upstream 3′ss signal.

It has been noted before that normal 5′ss sequences, in addition to sequences found inside certain exons, can act as splicing enhancers for the flanking intron upstream (20, 39, 48). Our data strongly suggest this is due to the ability of normal 5′ss sequences to pair with U1. In addition, our results demonstrate that pairing of U1 with other sequences in the area, inside the exon (7) or in the downstream intron, is sufficient to generate strong enhancer activity. In other words, binding sites for U1 can function as splicing enhancers both from within exons and from within introns. These observations support the suggestion that the usual exonic splicing enhancers provide an alternative means to recruit U1 when base pairing is not an option, for example, when a 5′ss signal is missing, as in 3′-terminal exons (2, 48).

**Dual effects of 5′ss mutations.** Mutations in the 5′ss sequence have two effects. First, they abolish splicing of the upstream intron, and this defect is consistently suppressed by coexpression of U1-α+14, indicating it is caused by disruption of the normal base pairing interaction between endogenous U1 and the 5′ss sequence. This suggests that the A5, CTC, and C1 mutations all abolish base pairing between endogenous U1 and the 5′ss sequence. Of course, it is also the basis of our suggestion above that the normal 5′ss sequence functions as a splicing enhancer and that base pairing of the sequence with endogenous U1 is necessary for this function. Second, the same mutations may or may not prevent correct splicing at the end of IDX. Presumably, this reflects the different roles individual nucleotides of the 5′ss consensus sequence play during recognition of the 5′ss or the subsequent catalytic steps. For example, mutations at the first intron position are known to block the second step of splicing (1, 13), and it was not unexpected, therefore, that U1-α+14 did not rescue intron C2 splicing for our position 1 mutant (C1). Similarly, when the 5′ss sequence had the CTC mutation, intron C2 splicing was not efficiently restored by U1-α+14, indicating that this mutation also interferes with proper 5′ss functioning. In contrast, splicing of intron C2 was observed for the A5 mutant in the presence of U1-α+14, which indicates that the consensus G at position 5 of intron C2 and base pairing between position 5 and U1 are not essential for 5′ss recognition or use, at least in this particular case. Thus, by using shift-U1 RNAs, the two functions of the 5′ss signal can be separated and the importance of specific nucleotides for 5′ss function, i.e., splicing of the downstream intron, determined without regard to their potential role in generating enhancer activity for the upstream intron.

When intron C2 splicing is not or incompletely restored in the presence of U1-α+14, there is no a priori reason to exclude the possibility that the remaining defect is caused by the loss of specific base pairs between the 5′ss sequence and U1. We have tested this possibility for many of our mutants, including A2, C3, and TCTC, using U1 RNAs with complete complementarity to the respective 5′ss sequences (collectively referred to as U1-α5′ss), but we have seen no substantial differences with U1-α+14 (Fig. 6B and 7 and unpublished data). In some cases, we have also determined whether U1-α5′ss improved intron C2 splicing when the upstream polypyrimidine tract was improved, but again, no large effects were observed (Fig. 6B and 7). This indicates that defects in 5′ss use are generally not due merely to a loss of base pairing with U1, but largely, if not only, to disruption of other interactions. This was confirmed for the CTC mutation by our experiment with U6-αCTC, which was necessary for splicing of intron C2, while U1-αCTC was not when the requirement for enhancer activity was abolished (Fig. 7). For the TCTC mutation, we have found that restored complementarity between the mutated 5′ss sequence and both U1 and U6 did not rescue intron C2 splicing (unpublished results), suggesting that the defect in this case may be due to the loss of some other function normally supported by position 3.

**Base pairing between U1 and the 5′ss sequence may be dispensable for accurate splicing of the downstream intron.** Although we suggest elsewhere that strong base pairing between U1 RNA and the 5′ss sequence can increase splicing of the downstream intron (17), it is clear that the downstream intron can be spliced when the complementarity between U1 and the 5′ss sequence is not sufficient to generate enhancer activity for the upstream intron. Single-base changes like A5 greatly reduce or abolish splicing of the upstream intron (Fig. 5A, but see also Fig. 2A), indicating they prevent base pairing of the 5′ss sequence with U1, but it is possible that base pairs can form at the remaining five positions of complementarity between the two sequences and that this is necessary for intron C2 splicing. However, when the complementarity between endogenous U1 and the 5′ss sequence was further reduced to three positions, intron C2 could still be spliced without increasing the complementarity between U1 and the 5′ss sequence (Fig. 7). This raises the question whether base pairing between U1 and the 5′ss sequence is necessary at all for splicing of the downstream intron. Of course, a definitive answer cannot be obtained in vivo. It seems unlikely, however, that a three-base complementarity is sufficient for initial selection of 5′ss on the basis of complementarity to U1 alone, but it is possible that SR proteins, believed to be involved in the initial recognition of 5′ss (9, 19, 39, 42, 54), are sufficiently selective in their interactions with pre-mRNA to direct base pairing between U1 and the 5′ss sequence when there are just three positions of complementarity. On the other hand, it is equally possible that no base pair formation is required, which would be in line with previous results demonstrating that splicing in vitro can proceed in the absence of U1 RNA (9, 42). Either way, our results show that substantial complementarity between U1 and the 5′ss sequence is important for splicing of the upstream intron but is not necessary for splicing of the downstream intron.

**Base pairing with U1 and exon definition.** The previous observation that U1 snRNA paired with the 5′ss sequence promotes binding of $U2AF^{65}$ to a weak polypyrimidine tract in the preceding intron suggested a mechanism for communication between splicing signals across exons (15), a key postulate

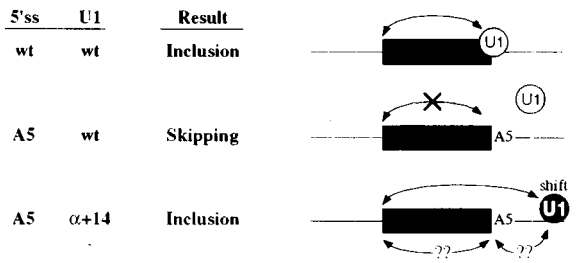



FIG. 8. Pairwise selection of splice sites without base pairing of U1 RNA with the 5′ss sequence. The first two lines illustrate the role of base pairing between U1 and the 5′ss sequence in communication between splice sites across the exon. An example of each situation is provided at the left. The third line illustrates a situation where U1 is paired at a distance from the 5′ss. Pairwise selection of splice sites must involve either direct communication between the sites, separate from communication between shift-U1 RNA and the 3′ss, or a connection between shift-U1 and the 5′ss. Our work argues against but does not formally exclude involvement of a second U1 molecule.

of the exon definition model (30). Of course, U2AF$^{65}$ does not directly determine the position of the 3′ss, but it assists in selection of the branch point(s), which may be the critical step in determining the exon's upstream boundary (11, 25, 31, 37). However, our data show that the position of the 5′ss, i.e., the exon's downstream boundary, is not determined relative to the position where U1 snRNAs that enhance splicing of the upstream intron pair with pre-RNA. For example, we have previously used a shift-U1 RNA, U1-α+12, with complementarity to the sequence at the cryptic 5′ss 12 bases into intron C2 (7); this is the site used to produce the cryp+12 protein (Fig. 2A and text). When coexpressed with the A5 reporter, U1-α+12 restored intron C1 splicing, indicating there was communication across the exon, but splicing at the end of IDX occurred at the normal position, not at the cryptic site (7 and unpublished data). Thus, coordinate definition of exon termini must involve either a second set of exon-spanning interactions or communication between U1 and the factors that ultimately determine the position of 5′ cleavage, such as the U5 and U6 snRNPs or some 5′ss-specific protein (Fig. 8). Obviously, the first possibility is complex, because it would require integration of two separate types of exon-spanning interactions. The second possibility is much simpler, although it would require that U1 can connect from different positions on the surrounding pre-mRNA with factors acting at the 5′ss, since shift-U1 RNAs with complementarity to different exon or intron sequences can restore correct splicing. Experimental evidence presented elsewhere is consistent with a connection between U1 and U6 (17), and this could be important for exon definition. We note, however, that exon definition may not be a necessary step when the upstream intron has a strong 3′ss signal, and as in our gene, both introns are short. In this situation (Fig. 7, lane 7), the suggested connection between U1 and U6 may no longer be relevant and it is possible that splicing can proceed in this case without U1–pre-mRNA pairing at or near the 5′ss.

## ACKNOWLEDGMENTS

We thank Christine Milcarek for critical reading of the manuscript, Diana Martik and Jon Snow for general technical support, and John Cardamone for oligonucleotide synthesis.

This work was supported in part by grant CN-74 from the American Cancer Society.

## REFERENCES

1. **Aebi, M., H. Hornig, and C. Weissmann.** 1987. 5′ cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5′ splice region, not by the conserved 5′ GU. Cell **50:**237–246.
2. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. **270:**2411–2414.
3. **Cáceres, J. F., S. Stamm, D. M. Helfman, and A. R. Krainer.** 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science **265:**1706–1709.
4. **Cohen, J. B., S. D. Broz, and A. D. Levinson.** 1989. Expression of the H-*ras* proto-oncogene is controlled by alternative splicing. Cell **58:**461–472.
5. **Cohen, J. B., S. D. Broz, and A. D. Levinson.** 1993. U1 small nuclear RNAs with altered specificity can be stably expressed in mammalian cells and promote permanent changes in pre-mRNA splicing. Mol. Cell. Biol. **13:** 2666–2676.
6. **Cohen, J. B., and A. D. Levinson.** 1988. A point mutation in the last intron responsible for increased expression and transforming activity of the c-Ha-*ras* oncogene. Nature (London) **334:**119–124.
7. **Cohen, J. B., J. E. Snow, S. D. Spencer, and A. D. Levinson.** 1994. Suppression of mammalian 5′ splice site defects by U1 small nuclear RNAs from a distance. Proc. Natl. Acad. Sci. USA **91:**10470–10474.
8. **Cortes, J. J., E. J. Sontheimer, S. D. Seiwert, and J. A. Steitz.** 1993. Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites *in vivo*. EMBO J. **12:**5181–5189.
9. **Crispino, J. D., B. J. Blencowe, and P. A. Sharp.** 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science **265:**1866–1869.
10. **Dominski, Z., and R. Kole.** 1991. Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. **11:**6075–6083.
11. **Gaur, R. K., J. Valcárcel, and M. R. Green.** 1995. Sequential recognition of the pre-mRNA branch point by U2AF$^{65}$ and a novel spliceosome-associated 28-kDa protein. RNA **1:**407–417.
12. **Ge, H., and J. L. Manley.** 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell **62:**25–34.
13. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7:**559–599.
14. **Hampson, R. K., L. La Follette, and F. M. Rottman.** 1989. Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. Mol. Cell. Biol. **9:**1604–1610.
15. **Hoffman, B. E., and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon. Genes Dev. **6:**2554–2568.
16. **Hosijima, K., K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura.** 1991. Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. Science **252:**833–836.
17. **Hwang, D.-Y., and J. B. Cohen.** 1996. U1 snRNA promotes the selection of nearby 5′ splice sites by U6 snRNA in mammalian cells. Genes Dev. **10:**338–350.
18. **Kandels-Lewis, S., and B. Séraphin.** 1993. Role of U6 snRNA in 5′ splice site selection. Science **262:**2035–2039.
19. **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 (London) **368:** 119–124.
20. **Kuo, H., F. H. Nasim, and P. J. Grabowski.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science **251:**1045–1050.
21. **Lavigueur, A., H. La Branche, A. R. Kornblihtt, and B. Chabot.** 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. **7:**2405–2417.
22. **Lesser, C. M., and C. Guthrie.** 1993. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science **262:**1982–1988.
23. **Madhani, H. D., and C. Guthrie.** 1994. Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet. **28:**1–26.
24. **Mayeda, A., D. M. Helfman, and A. R. Krainer.** 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol **13:**2993–3001.
25. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to mRNA by the spliceosome, p. 303–357. *In* R. F. Gesteland and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
26. **Nagoshi, R. N., and B. S. Baker.** 1990. Regulation of sex-specific RNA splicing at the *Drosophila doublesex* gene: *cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev. **4:**89–97.
27. **Newman, A. J., and C. Norman.** 1991. Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. Cell **65:**115–123.
28. **Newman, A. J., and C. Norman.** 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell **68:**743–754.
29. **Ramchatesingh, J., A. M. Zahler, K. M. Neugebauer, M. B. Roth, and T. A. Cooper.** 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol. Cell. Biol. **15:**4898–4907.
30. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10:**84–94.
31. **Rosbash, M., and B. Séraphin.** 1991. Who's on first? The U1 snRNP-5′

splice site interaction and splicing. Trends Biochem. Sci. **16:**187–190.
32. **Rymond, B. C., and M. Rosbash.** 1992. Yeast pre-mRNA splicing, p. 143–192. *In* E. W. Jones, J. R. Pringle, and J. R. Broach (ed.), The molecular and cellular biology of the yeast *Saccharomyces*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
33. **Sawa, H., and Y. Shimura.** 1992. Association of U6 snRNA with the 5′-splice site region of pre-mRNA in the spliceosome. Genes Dev. **6:**244–254.
34. **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.
35. **Séraphin, B., and M. Rosbash.** 1990. Exon mutations uncouple 5′ splice site selection from U1 snRNA pairing. Cell **63:**619–629.
36. **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.
37. **Smith, C. W. J., E. B. Porro, J. G. Patton, and B. Nadal-Ginard.** 1989. Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. Nature (London) **342:**243–247.
38. **Sontheimer, E. J., and J. A. Steitz.** 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science **262:**1989–1996.
39. **Staknis, D., and R. Reed.** 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol. **14:**7670–7682.
40. **Strijker, R., D. T. Fritz, and A. D. Levinson.** 1989. Adenovirus VAI-RNA regulates gene expression by controlling stability of ribosome-bound RNAs. EMBO J. **8:**2669–2675.
41. **Sun, Q., A. Mayeda, R. K. Hampson, A. R. Krainer, and F. M. Rottman.** 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev. **7:**2598–2608.
42. **Tarn, W.-Y., and J. A. Steitz.** 1994. SR proteins can compensate for the loss of U1 snRNP functions in vitro. Genes Dev. **8:**2704–2717.
43. **Tian, M., and T. Maniatis.** 1992. Positive control of pre-mRNA splicing *in vitro*. Science **256:**237–240.
44. **Tian, M., and T. Maniatis.** 1993. A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. Cell **74:**105–114.
45. **van Oers, C. C. M., G. J. Adema, H. Zandberg, T. C. Moen, and P. D. Baas.** 1994. Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol. Cell. Biol. **14:**951–960.
46. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.
47. **Wassarman, D. A., and J. A. Steitz.** 1992. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science **257:**1918–1925.
48. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7:**407–418.
49. **Woolford, J. L.** 1989. Nuclear pre-mRNA splicing in yeast. Yeast **5:**439–457.
50. **Wu, J. Y., and T. Maniatis.** 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75:**1061–1070.
51. **Wyatt, J. R., E. J. Sontheimer, and J. A. Steitz.** 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. Genes Dev. **6:**2542–2553.
52. **Xu, R., J. Teng, and T. A. Cooper.** 1993. The cardiac troponin T alternative exon contains a novel purine-rich splicing element. Mol. Cell. Biol. **13:**3660–3674.
53. **Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. **6:**837–847.
54. **Zuo, P., and J. L. Manley.** 1994. The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites. Proc. Natl. Acad. Sci. USA **91:**3363–3367.