MOLECULAR AND CELLULAR BIOLOGY, June 1990, p. 2874–2881
0270-7306/90/062874-08$02.00/0

Vol. 10, No. 6

U1 Small Nuclear RNA from *Schizosaccharomyces pombe* Has Unique and Conserved Features and Is Encoded by an Essential Single-Copy Gene

GREGORY PORTER, PATRICK BRENNWALD, AND JO ANN WISE*
Department of Biochemistry, University of Illinois, Urbana, Illinois 61801

Received 3 January 1990/Accepted 5 March 1990

We have cloned, sequenced, and disrupted the gene encoding U1 small nuclear RNA (snRNA) in the fission yeast *Schizosaccharomyces pombe*. This RNA is close in size and exhibits a high degree of secondary structure homology to human U1 RNA. There exist two regions of extended primary sequence identity between *S. pombe* and human U1 RNAs; the first comprises nucleotides involved in hydrogen bonding to 5′ splice junctions, and the second is a single-stranded region which, in the human snRNA, forms part of the A protein binding site. *S. pombe* U1 lacks two nucleotides just following the 5′ cap structure which are present in all other U1 homologs examined to date, and the region which corresponds to the binding site for the human 70K protein (molecular weight of 55,000) is more divergent than in other organisms. A putative upstream transcription signal is conserved in sequence and location among all loci encoding spliceosomal snRNAs in *S. pombe* with the exception of U6. Disruption of the single-copy U1 gene, designated *snu1*, reveals that this RNA is indispensable for viability.

The nuclei of all eucaryotic cells contain a class of low-molecular-weight capped RNA species designated U-class small nuclear RNAs (snRNAs) (for a review, see reference 12). Five of these (U1, U2, U4, U5, and U6) are components of small nuclear ribonucleoprotein particles (snRNPs) that participate in the splicing of mRNA precursors (for a review, see reference 44). A primary function of the U snRNPs is to recognize splicing signals in the pre-mRNA. Even before the mechanism of intron removal was understood, binding of U1 to the substrate was proposed on the basis of potential hydrogen bonding between nine nucleotides near the 5′ terminus of the snRNA and mammalian 5′ splice junctions. Analysis of compensatory base changes in both human (50) and *Saccharomyces cerevisiae* (budding yeast) cells (37, 38, 40) provided direct proof of base pairing between U1 and 5′ splice sites at some, but not all, positions tested. The first 10 nucleotides of U1 snRNA from *S. cerevisiae* and mammals are identical in spite of differences in the 5′ splice site consensus sequences, suggesting that they may have function(s) in addition to substrate recognition (12). Yeast U1 is nearly four times as large as human U1 (18, 41), yet it lacks certain structures otherwise universally conserved (17).

The protein components of snRNPs include a common core recognized by specific autoimmune sera, as well as a variable number of polypeptides uniquely associated with a particular snRNA (for a review, see reference 21). The recent isolation of cDNA clones for the three U1 snRNP-specific proteins has allowed mapping of their binding sites on the RNA. A variety of in vitro assays demonstrate that the 70K protein (molecular weight of 55,000) interacts specifically with stem-loop I (31, 34, 45), while the A protein binds to stem-loop II (22, 30). The C protein also appears to be associated with the 5′ portion of U1 (14), but some controversy exists as to its specific binding site on the RNA (30).

* Corresponding author.

Signals which direct the synthesis of U snRNAs have been extensively characterized in higher eucaryotes (for a review, see reference 9). On the basis of α-amanitin sensitivity, U snRNAs are thought to be transcribed by RNA polymerase II, with the exception of U6, which is a pol III transcript. Several features distinguish their synthesis from that of mRNAs, however. Metazoan snRNA promoters lack a canonical TATA sequence; instead, the position of transcription initiation is specified by a conserved motif centered at −50 to −60 (versus −25 to −35 for mRNA TATA boxes). A distal sequence element located between −200 and −250 is similar to, but not interchangeable with, enhancers located upstream from mRNA coding sequences. Perhaps most intriguing is the 3′ box located just downstream of the human U1 and U2 genes, which directs formation of the proper 3′ end but is bypassed if transcription was initiated at an mRNA rather than an snRNA promoter. Little is known about snRNA transcription in lower eucaryotes but, in contrast to the situation in mammals, a construct in which synthesis of the *S. cerevisiae* U5 homolog was driven by an mRNA promoter allowed accurate 3′ end formation (29).

Nuclear pre-mRNA splicing has been studied most extensively in mammalian and *S. cerevisiae* cells. Although introns are removed by similar mechanisms in both systems, detailed analysis has revealed differences in, for example, the structures of the snRNAs and the effects of mutating the signals which they recognize in the pre-mRNAs. To elucidate which aspects of splicing are universal and which are species specific, we have turned our attention to the fission yeast *Schizosaccharomyces pombe*, whose phylogenetic position is nearly as distant from *S. cerevisiae* as from humans (36). We previously described homologs of the spliceosomal RNAs U2 (7) and U5 (43), as well as U3 snRNA (32), which is though to be involved in ribosome biogenesis. Other laboratories have cloned and sequenced the genes encoding U4 (11) and U6 (46). Here we report the cloning, structural analysis, and disruption of a gene encoding U1 from *S. pombe*, thus completing the roster of spliceosomal snRNAs in this organism. This RNA is the smallest U1 homolog

sequenced to date, yet it exhibits all the features previously demonstrated experimentally or predicted on the basis of evolutionary conservation to be important for function. We also report the results of aligning the DNA flanking each snRNA coding region to identify conserved sequences; a proximal element which includes a TATA motif is located in a similar position upstream from all *S. pombe* snRNA genes except U6, but no extensive conservation is found in the downstream DNA. Gene disruption reveals that, as expected, U1 snRNA is essential for viability in *S. pombe*.

MATERIALS AND METHODS

Chemicals and enzymes. Restriction enzymes, polynucleotide kinase, and DNA ligase were purchased from Bethesda Research Laboratories, Inc. T7 DNA polymerase (Sequenase) and other DNA sequencing reagents were purchased from U.S. Biochemical Corp. Reverse transcriptase was purchased from Life Sciences, Inc. Enzymatic RNA sequencing reagents were purchased from Pharmacia, Inc. DNA polymerase and other nick translation reagents were purchased from Amersham Corp. Radiochemicals were obtained from ICN Pharmaceuticals Inc.

Strains and cloning vectors. *S. pombe* haploid strain 972 (h^-) was propagated as described by Gutz et al. (13). The SP629 strain used in the gene disruption (see below) was maintained as a diploid by constant selection for adenine prototrophy resulting from intergenic complementation of two *ade6* alleles. The U1 gene was isolated from a genomic bank constructed by Losson and Lacroute in the plasmid pFL20 (20), propagated in *Escherichia coli* LE392. Fragments used for hybridization, sequence analysis, and construction of the gene disruption were subcloned into the vector pUC19 and propagated in *E. coli* TG1. Fragments used for gene complementation were subcloned into the shuttle vector pIRT2 (described in reference 19), in which the *LEU2* gene was replaced by *S. cerevisiae URA3*. Recombinant DNA manipulations were as described by Maniatis et al. (23) unless otherwise noted.

Nucleic acid preparation and analysis. Genomic DNA and total RNA from *S. pombe* were extracted as previously described (6). Oligonucleotides were synthesized at the University of Illinois Biotechnology Center. Those used for DNA sequencing and/or primer extension on RNA are U1B, 5′-TAAATATGGGTGCTT-3′; U1PE, 5′-TCAGTGCAATG CCA-3′; and Bb15, 5′-CCTCATTTGGGGCAA-3′. Their locations are indicated in Fig. 2. Protocols for DNA and RNA sequence analysis and primer extension were described previously (6, 7). Conditions for colony hybridization and Southern analysis have also been described previously (6).

Construction and analysis of a gene disruption. To create a null allele of *snu1*, a 940 nucleotide (nt) *Hin*cII fragment carrying the gene was replaced with the *LEU2* gene from *S. cerevisiae* carried on a *Hin*dIII fragment rendered blunt ended by treatment with the Klenow fragment. The plasmid, cleaved with *Pvu*II and *Hin*dIII to release a linear fragment containing the disrupted gene, was transformed into *S. pombe* SP629, a kind gift of David Beach (genotype h^+/h^- *ade6-210/ade6-216 leu1-32/leu1-32 ura4-d18/ura4-d18*); the *LEU2* gene can complement mutations in *S. pombe leu1*. Transformants were selected for leucine prototrophy on EMM2 plates (25) containing 1 M sorbitol and 100 mg of uracil per liter. After streaking for single colonies on EMM2 plus uracil, cells containing stably integrated *LEU2* were identified as those which retained the marker following ca. 15 generations of growth in EMM2 liquid medium supplemented with 100 mg of leucine per liter in addition to uracil. Cells were plated from the liquid suspensions onto rich medium (YEA; 13) and diploid (white) colonies were chosen for further analysis (haploids are red due to adenine deficiency). Of 35 Leu^+ colonies initially chosen, 7 were stable transformants; 5 of these appeared to carry the desired gene disruption, as monitored by Southern blot analysis (see Results). One strain, *snu1ΔLEU15*, was selected for tetrad and random spore analysis, performed as previously described (6, 19).

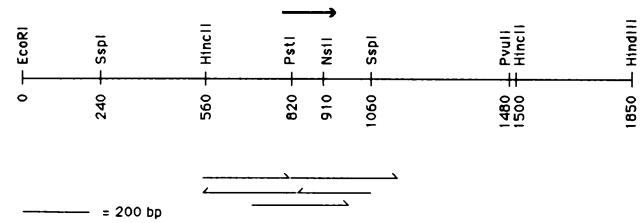

FIG. 1. Restriction map of the 1.85-kb *Eco*RI-*Hin*dIII fragment containing the U1 snRNA gene. Restriction mapping was performed on this fragment after subcloning it into pUC19 to yield the plasmid pBHE. The location of the mature U1 transcript is indicated by a thick arrow above the map; thin arrows below the map illustrate the strategy used to sequence the *snu1* gene and flanking DNA.

RESULTS

Structure and organization of the U1 snRNA gene. A U1 snRNA candidate was initially identified on the basis of similarity in size to human U1 on a two-dimensional gel of total RNA from *S. pombe* (6). Screening of a genomic library (20) with a radioactive cDNA probe identified a single hybridizing plasmid, designated pBb, containing a 3.4-kilobase (kb) insert. The RNA coding region was located by Southern blot hybridization with a cDNA probe, and a 1.85-kb *Eco*RI-*Hin*dIII subclone was generated for detailed analysis. Figure 1 shows the restriction map of this fragment; the location and orientation of the gene and the extent and direction of the DNA sequence determined are indicated. Recently, a putative U1 gene was identified by probing an *S. pombe* library with a human clone (10); the *Eco*RI-*Hin*dIII fragment reported to hybridize to the human gene is similar in size to the one we mapped in detail.

Figure 2 shows the primary structure of the *S. pombe* U1 gene, designated *snu1*, together with 239 nucleotides of upstream sequence and 120 nucleotides of downstream sequence. The 5′ terminus of the RNA was determined by primer extension analysis (Fig. 3) using a U1-specific oligonucleotide on total RNA. The 3′ terminus of the RNA was determined by direct RNA sequencing (data not shown) of [5′-^{32}P]pCp-labeled U1 isolated from a gel following precipitation by antibodies directed against trimethylguanosine, as described previously (7). Its reactivity with the anticap serum demonstrates that *S. pombe* U1 possesses the 5′ end structure characteristic of U-class snRNAs.

Sequence and structure comparison of *S. pombe* and human U1 RNAs. An alignment of U1 snRNA sequences which includes our *S. pombe* U1 data has been published previously in a review article (12). This RNA can fold into a secondary structure virtually identical to that of human U1 (Fig. 4). The moderately smaller size of the *S. pombe* RNA is primarily due to a shorter single-stranded region surrounding the Sm site and a truncated stem-loop IV. As is generally true for functional RNAs, regions of primary sequence

```
      -230       -220       -210       -200       -190       -180       -170       -160
GTTAACTATT CATGGCTTGA AAGCAACGAA ATCATTTCCT ATTATGAAAT TTTGGAGATT TATATAATTA CAAATCGCTG

      -150       -140       -130       -120       -110       -100        -90        -80
AACGATTTCAC TAATCAGCA AAATTTAACG ATTTAAATAT CATTTTTGTT AAATAATTAT ACTGTACTTC TTTATTGCAT

       -70        -60        -50        -40        -30        -20        -10         -1
GTAAGAATGT GACTACTTTC GTGACTGCTT CGTTCCGCGT CGTTACTATA AATATGGGTG CTTACACTTA CGGTTCTTT
                                                      U1B

          +1         10         20         30         40         50         60         70
           ACTTACCTGG CATGAGTTTC TGCAGCACAA GAATTGTGGA GACTCAGTTA TTTGTCTTGG CATTGCACTG
                                                                          U1PE
m2,2,7GpppACUUACCUGG CAUGAGUUUC UGCAGCACAA GAAUUGUGGA GACUCAGUUA UUUGUCUUGG CAUUGCACUG

        80         90        100        110        120        130        140       148
AGCCCTGACG AATAACTGTG GACTGGCTAA GGTCAGCTCC GGATGCATCA TTTTTGAGTT CGTCCCTCAT TTGGGGCAAT
                                                                          Bb15
AGCCCUGACG AAUAACUGUG GACUGGCUAA GGUCAGCUCC GGAUGCAUCA UUUUUGAGUU CGUCCCUCAU UUGGGGCAOH

       +12        +22        +32        +42        +52        +62        +72        +82
TTTACTAATT AAGGCGATAT GGTTGCCTTA AATAGTAAAG TTATCCGAAA TGGATTATGT AAAGTAACAT TTTCTTTCAT

       +92       +102       +112      +120
CATATATAAT TGTAAATTAT TGTATCATAA TTAATATT
```

FIG. 2. Nucleotide sequence of the U1 gene and flanking DNA. The sequence of the nontranscribed strand (confirmed on the transcribed strand) is shown on the top line; the U1 RNA sequence, derived in part from the DNA sequence and in part from direct analysis of the RNA (see the text), is shown beneath the coding region. The RNA is shown as capped with $m^{2,2,7}G$ because it is precipitable with antibodies directed against this structure. Underlined DNA sequences denote oligonucleotides used as sequencing primers. The sequence starting from −239 and proceeding toward the gene was confirmed by using the universal primer on a pUC19 subclone of a *Hin*cII-*Pst*I fragment (Fig. 1). The downstream flanking sequence was confirmed by sequencing from the universal primer on an *Ssp*I subclone. The GenBank accession number for this sequence is M29062.

identity between *S. pombe* and human U1 are predominantly single stranded and correspond to sites of known or suspected interactions (see Discussion). The 5′ end is highly conserved, as in other U1 RNAs; however, the *S. pombe* U1 RNA lacks two nucleotides just following the cap structure

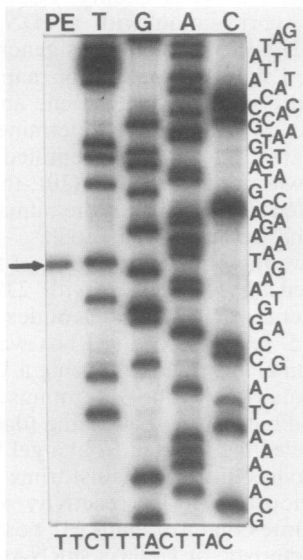

FIG. 3. Identification of the 5′ end of *S. pombe* U1 snRNA. To determine the start site of transcription, we performed primer extension (PE) using an oligonucleotide complementary to nucleotides 57 to 71 of the RNA, designated U1PE in Fig. 2. A DNA sequencing ladder (T, G, A, C) obtained from the same primer was run as a control. The primer extension product is indicated by an arrow. The sequence written alongside the gel is that of the transcribed strand; the sequence of the nontranscribed strand is shown below the gel in 5′→3′ orientation, with the initiating nucleotide underlined.

which are present in all previously identified homologs, including that of *S. cerevisiae* (18, 41). Loop I of *S. pombe* U1, although clearly related to the corresponding sequence in the human RNA, is the most divergent example sequenced to date. The first eight nucleotides of loop II, which are well conserved in metazoan U1 RNA (12), are identical in the human and *S. pombe* RNAs. Stems III and IV are present in *S. pombe* U1, but the loops at the ends of these helices lack extensive sequence identity with the human RNA. All of the hairpins in both human and *S. pombe* U1 contain internal loops or bulged nucleotides, with the exception of stem I in the latter RNA; however, only the bulged C in stem III is conserved in both identity and position. The Sm site in fission yeast U1, presumably responsible for binding to the common snRNP protein core, conforms to the canonical sequence $AU_{4-6}G$ (5). This contrasts with U2, U4, and U5 snRNAs from this organism, which all have a G preceding the run of uridines (7, 11, 43).

Identification of a putative transcription signal upstream from *S. pombe* snRNA genes. With the present work, sequence data extending through the regions essential for synthesis of higher eucaryotic snRNAs are now available for all the major *S. pombe* U-class snRNA loci. In order to determine whether these flanking sequences contain conserved elements which might be functionally analogous to transcription signals characterized in metazoan snRNA genes (for a review, see reference 9) and recently identified in plants (47), we performed a computer-assisted alignment. Figure 5 shows a region of extensive sequence similarity centered at 28 to 33 nt upstream of the start site that is shared by all *S. pombe* snRNA genes with the exception of U6. Other identities (not shown) included an additional 12 of 12 match located further upstream from the U1 and U2 coding sequences which was not found at the other loci, and two larger but imperfect tracts conserved only between U1 and U3. We were unable to identify even a minimal degree of sequence conservation downstream from *S. pombe* snRNA

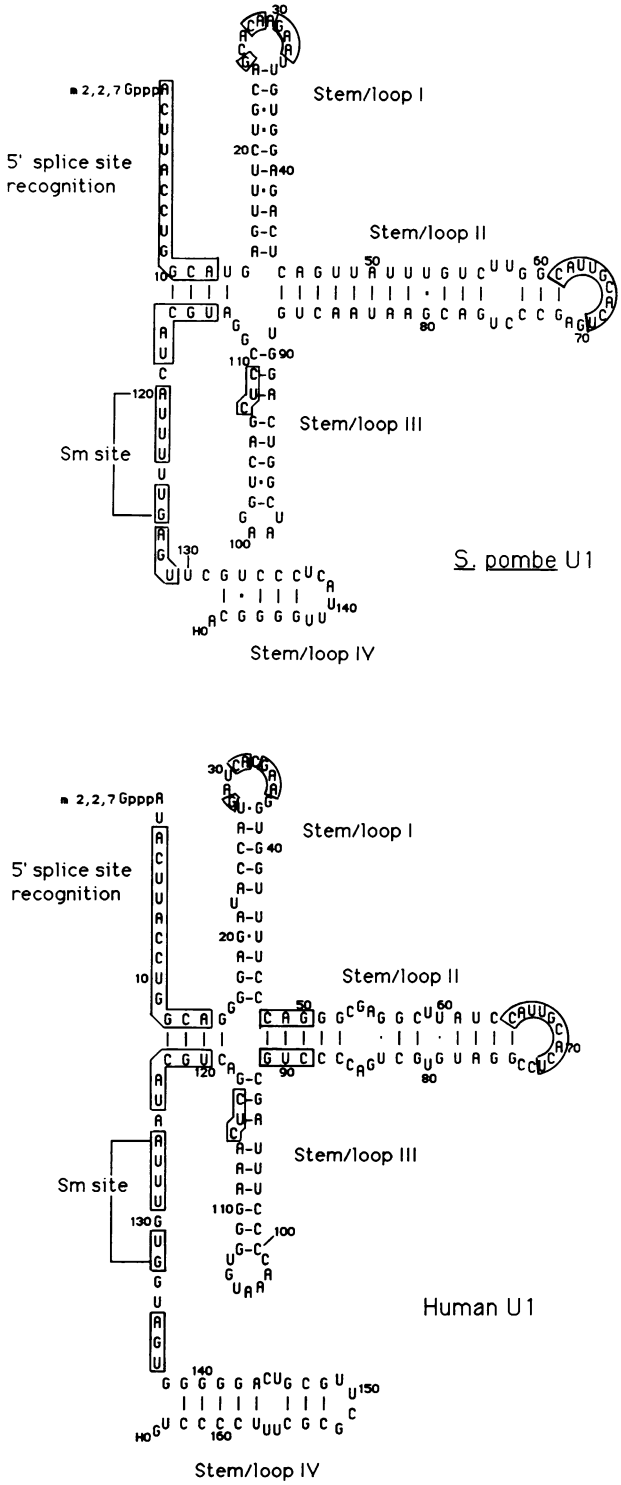

FIG. 4. Primary and secondary structure comparison of *S. pombe* and human U1 snRNAs. The folding pattern for *S. pombe* U1 was derived by phylogenetic comparison with previously identified homologs. The secondary structure of human U1 is adapted from that described by Branlant et al. (4) and is based on a combination of phylogenetic and structure-probing data. Sequence identities located in homologous regions in the secondary structures are indicated by boxes. Nomenclature of the structural elements is as described by Guthrie and Patterson (12).

genes; the limited identity in the 3'-flanking DNA of U2, U3, and U4 previously reported (11) is not supported when the U1 and U5 sequences are included in the sample.

***S. pombe* U1 is encoded by an essential single-copy gene.** Since spliceosomal snRNAs have been shown to be essential in *S. cerevisiae*, we set out to determine whether the same was true for *S. pombe* U1. A null allele was created in vitro (Fig. 6*a*) and introduced into a diploid *S. pombe* strain by transformation (see Materials and Methods). Figure 6*b* shows a Southern analysis which confirms that integration occurred at the correct locus. Upon digestion with *Eco*RI, a single 3.4-kb band was observed in the parental strain, while the diploid heterozygous for *snu1* gene disruption showed this band and two additional hybridizing fragments at 1.9 and 2.7 kb. The former extends from the *Eco*RI site upstream of the U1 coding sequence to the *Eco*RI site within the *LEU2* gene; the latter is comprised of the remainder of the *LEU2* gene and the U1 3'-flanking sequence extending to the *Eco*RI site. Digestion with *Hin*dIII and *Xba*I reveals a single 3.6-kb band in the parental DNA; the additional band in the transformed diploid extends from the *Xba*I site upstream of the U1 coding sequence to the *Hin*dIII site in the 3'-flanking DNA.

If an essential gene has been disrupted, upon sporulation, 2:2 segregation for viability should be observed in tetrads, and no viable progeny should be Leu^+. These predictions were borne out by dissection of 17 asci, in which never more than two viable spores were observed, all being Leu^- (data not shown). Tetrads with four viable spores were obtained from the parental diploid. The indispensability of an intact U1 locus was further confirmed by analysis of 120 random spores; again, no Leu^+ progeny were observed. The inviability of these cells is due solely to deletion of the U1 coding sequence, as demonstrated by complementation experiments in which the gene and flanking sequences were introduced on a multicopy plasmid (see Materials and Methods). Both constructions tested, a 1,480-base-pair (bp) *Eco*RI-*Pvu*II fragment containing ca. 800 nt of 5'- and 520 nt of 3'-flanking sequence and a 500-bp *Hin*cII-*Ssp*I fragment containing 239 nt of 5'- and 120 nt of 3'-flanking sequence (Fig. 1), could rescue spores carrying the gene disruption. Ura^+ Leu^+ and Ura^+ Leu^- colonies segregated approximately 1:1 (data not shown).

DISCUSSION

The U1 snRNA-5' splice junction interaction: similarity between *S. pombe* and mammals. Figure 7 shows potential hydrogen bonding between the 5' splice junction consensus sequences for *S. pombe*, *S. cerevisiae*, and mammalian pre-mRNAs and their corresponding U1 RNAs. The 5' truncation of the *S. pombe* snRNA does not affect its ability to base pair with the splice site. In all three cases, any given precursor has 5 to 7 nucleotides of complementarity to U1 snRNA; in *S. pombe* and mammals, this is accomplished by the variation of individual 5' splice site sequences around the consensus, while in *S. cerevisiae* there is a programmed mismatch between U1 and an invariant 5' junction. Creation of a perfect hybrid with the splicing substrate through mutation of *S. cerevisiae* U1 resulted in a significantly longer generation time (38, 40), suggesting that too strong an interaction is deleterious.

Does the structural difference in the U1-precursor interaction underlie mechanistic differences between *S. cerevisiae* and mammalian splicing? Although a complete answer is not yet available, it is noteworthy that while the flexibility

```
                    -45                                          -10        +1
U1:   GCGACTGCTTCGTTCCGCGTCGTTACTATAAATAT-GGGTCGTTACACTTACGGATTCTTTACU
                  -45       ♦♦♦♦♦♦♦♦♦♦♦♦ ♦♦♦♦       -10        +1
U2:   GGGTTCGGTTTATTTATGAGATTTACTATATATAC-AGGTACTCGTGCATTTGGTTCGGTAUU
                   -45  ♦♦♦♦♦♦♦♦♦♦♦♦♦♦ ♦♦♦♦          -10        +1
U3:   GAATGTGACTCTTTCGTTATATTTACTATATATAC-GGGTCATTATCCTACGAACAAAACACTACG
                 -45       ♦♦♦♦♦♦♦♦♦♦♦♦♦♦ ♦♦♦♦       -10        +1
U4:   GACCTGTTGTGACTGCTCAATTTTACTATAAATGG-ATGTACAAGAGTTTACCATTCCCAUC
               -45         ♦♦♦♦♦♦♦♦♦♦♦♦  * ♦♦      -10        +1
U5:   TAGGGTTTGAGGTTGCAAAAAGTTA-TATAAACACCTGGTTGCACACACATTTCAGCAUA
                           ♦ ♦♦♦ ♦♦♦♦♦♦♦♦  ♦♦♦
CONS:                      atTTAcTATAWAYRY-RgGT
```

FIG. 5. Conserved sequences upstream from *S. pombe* snRNA coding regions. A computer-aided search for putative transcription signals was performed by using the DNA* software package. Numbering is relative to the start site of transcription (+1); the last digit of each number is aligned with the corresponding nucleotide. A consensus sequence (CONS) was derived for the only extended similarity found when the U1 to U5 snRNA flanking sequences were compared. Uppercase letters indicate conservation among all five loci; lowercase letters designate nucleotides conserved in the majority of cases. Abbreviations: N, any nucleotide; R, purine; Y, pyrimidine; W, A or T. The symbols below each sequence indicate identity to defined nucleotides within the consensus sequence (◆) or to degenerate positions (*). The sequences compared were as follow: for U1, 239 nt upstream and 120 nt downstream (this work, Fig. 2); for U2, 250 nt upstream and 91 nt downstream (P. Brennwald, K. Schaefer, H. Skinner, and J. A. Wise, unpublished data); for U3, 314 nt upstream and 62 nt downstream (32); for U4, 208 nt upstream and 463 nt downstream (11); and for U5, 250 nt upstream and 77 nt downstream (P. Brennwald, K. Small, H. Skinner, and J. A. Wise, unpublished data).

of mammalian splicing signals often allows the use of cryptic junctions when the normal one is mutated (for a review, see reference 28), this generally does not occur in *S. cerevisiae*. The strict conservation of 5′ splice junction sequences in *S. cerevisiae* may also contribute to its inability to splice mammalian pre-mRNAs, while *S. pombe* can do so in at least one case (see reference 15 and references therein). We are currently using mutational analysis to test the prediction that 5′ splice junction recognition in *S. pombe* resembles this process in mammals.

Variable conservation of U1 protein-binding sites. Stem-loop I has been shown to specifically interact with the human U1 snRNA-associated 70K protein in vitro (31, 34, 45). Although the stem is similar in length, loop I in *S. pombe* U1 is more divergent from the human U1 sequence than in any previously identified homolog (12). Mutating nucleotides in the highly conserved loop produces variable effects on human U1 function in vivo (49) which do not perfectly correlate with their evolutionary conservation. For example, similar, moderate decreases in splicing efficiency were observed in human cells upon mutating U1 position 29, which is not conserved in the *S. pombe* RNA, and positions 32 and 34 to 36, which are. However, the most severe phenotype resulted from mutating the C residue at position 31 of the human RNA, which is invariant in all U1 homologs, including that of *S. pombe*.

Stem-loop II binds to the human U1-specific A protein (22, 30); the length of this stem is similar, and the first eight nucleotides of the loop are identical, between the human and *S. pombe* RNAs, suggesting that *S. pombe* may contain a highly conserved homolog of this protein. In humans, the sequence of the A protein is related to that of the U2 snRNP-specific B″ protein (42), and stem-loop III of U2 has been proposed as a binding site for the B″ protein based in part on sequence similarity to stem-loop II of U1 (22). However, since the common bases are mostly paired in the U2 molecule but single stranded in U1, it seems unlikely that their mode of interaction with a protein would be analogous.

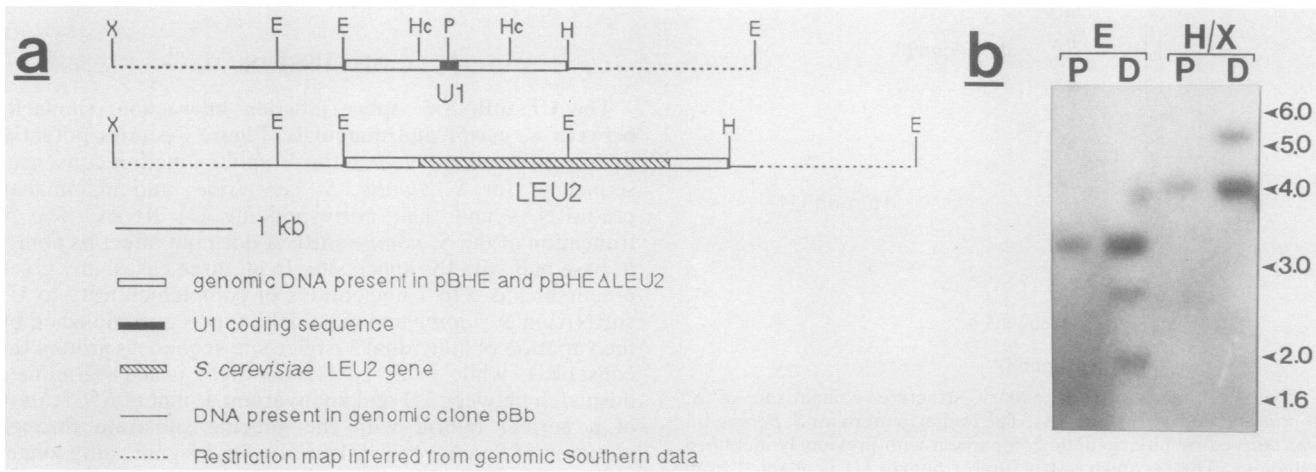

FIG. 6. Construction and analysis of a U1 gene disruption. (*a*) Restriction map of the gene disruption compared with that of the wild-type genomic locus (see the text for details of the construction). Abbreviations: X, *Xba*I; E, *Eco*RI; Hc, *Hin*cII; P, *Pst*I; H, *Hin*dIII. (*b*) Southern blot of genomic DNA from the parental diploid strain (SP629) and the strain heterozygous for disruption of the U1 gene (*snu1ΔLEU15*); the samples were digested with the indicated restriction enzymes, transferred to Hybond, and probed with nick-translated pBHE. Abbreviations: P, DNA from the parent; D, DNA from the disruptant; E, *Eco*RI; H, *Hin*dIII; X, *Xba*I. Size markers were the 1-kb ladder (Bethesda Research Laboratories).

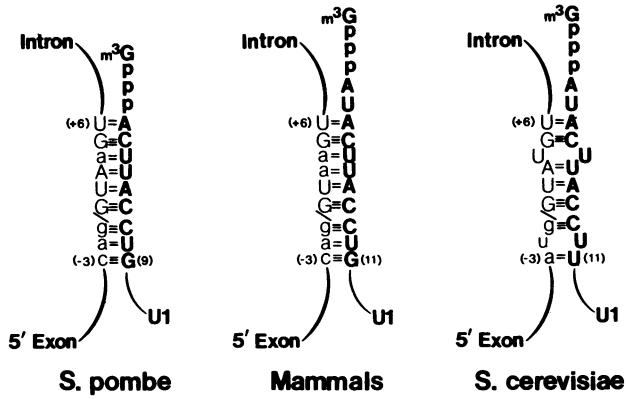

FIG. 7. Proposed base pairing of U1 snRNAs with 5′ splice junctions. In all three diagrams, the sequence of the 5′ end of the U1 RNA is shown in bold letters and the pre-mRNA 5′ splice junction consensus sequence is shown in fine print. The *S. pombe* consensus sequence is as described previously (15), supported by more recent sequence data (2, 3, 27); pairing patterns shown for human and *S. cerevisiae* U1 RNAs are adapted from those described by Siliciano et al. (41). For each 5′ splice site, uppercase letters indicate nucleotides which are greater than 80% conserved while lowercase letters denote less than 70% conservation.

We propose instead that B″ might bind to another, unpaired region of U2 which is similar in primary sequence to the A protein binding site on U1; the conservation between loop IV of U2 and loop II of U1 was first noted in an alga (16) and is preserved in other homologs, including those of *S. pombe* and humans (Fig. 4). Our proposal is consistent with earlier work implicating the 3′ half of U2 snRNA in the binding of proteins unique to this snRNP (24).

Bulged nucleotides are found in stems I and III of human U1 snRNA. A precedent for their importance was set by studies with a fragment of bacteriophage R17 RNA, which demonstrate that a bulged nucleotide within the helix is essential for coat protein binding (48). The bulged C in stem III of U1 RNA is a promising candidate for a specific interaction, since it is conserved in all homologs sequenced to date except those of plants, which have a bulged G in a similar position (12). Although a bulged nucleotide is found in stem I of metazoan U1 RNA, it has been shown to be dispensable for protein binding in vitro (34) and to have little effect on splicing efficiency in vivo (49); thus its absence from the *S. pombe* and *Kluyveromyces lactis* (17) RNAs is not surprising.

***S. pombe* and *S. cerevisiae* U1 RNAs are strikingly different.** *S. pombe* U1 is structurally homologous to human RNA over its entire length, in contrast to the *S. cerevisiae* homolog, in which only the 5′ single-stranded region and stem-loop I are clearly recognizable (18, 41). Despite its large size (568 nt), *S. cerevisiae* U1 RNA lacks structures with obvious homology to stem-loops II, III, and IV of metazoan U1 RNA, all of which are present in the *S. pombe* RNA, and instead contains several budding yeast-specific helices which are also found in the *K. lactis* homolog (17). A stem-loop structure following the Sm site, a feature found not only in U1 but in all U-class snRNAs which bind to this common core of proteins (5), is present in the *S. pombe* but not in the *S. cerevisiae* homolog. Only one aspect, the sequence of loop I, is more highly conserved in *S. cerevisiae* U1 RNA. These structural differences suggest that the interactions of U1 with snRNP proteins and with other spliceosome components may not be similar in the two yeasts.

Sequences resembling mammalian snRNA transcription signals are not found in the DNA flanking *S. pombe* snRNA genes. Since the locations and sequences of transcription signals for protein-coding genes appears to be similar between mammals and *S. pombe* (35), we compared the sequences flanking *S. pombe* snRNA genes to elements required for mammalian snRNA synthesis, as well as to each other. Our search revealed no similarity to metazoan snRNA transcription signals, and a single stretch of conserved nucleotides upstream from U1 to U5 genes in *S. pombe*. Although direct evidence that this element regulates transcription is lacking, we note that the amount of each snRNA strongly correlates with how well its upstream element conforms to the consensus sequence shown in Fig. 5 (see references 6 and 7 for experimental evidence to demonstrate this point). The conserved region includes a sequence related to the canonical TATA motif located in a similar position upstream from genes encoding mRNAs in most eucaryotes, including *S. pombe* (35). However, the remaining nucleotides are unique to the U1 to U5 genes. The *S. pombe* U6 gene has a TATA box in a similar position (46), but the surrounding sequences are not related to the conserved box upstream from the other *S. pombe* snRNA coding regions; the absence of an snRNA-specific upstream element implies that, as in mammals, U6 is transcribed by a different polymerase (9). TATA-rich sequences are likewise located in similar, but farther upstream, positions in both snRNA and protein-coding genes in *S. cerevisiae* (for an example, see reference 39). In contrast, genes encoding trimethylguanosine-capped snRNAs from higher eucaryotes are preceded by a proximal element which functions like the TATA box to specify the site of initiation, but is dissimilar in sequence (9). No convincing homology to the distal enhancer located between −200 and −250 of mammalian snRNA loci (9) was found preceding *S. pombe* snRNA genes. These differences are particularly surprising in light of the presence of sequences related to both the proximal and distal metazoan snRNA transcription signals upstream of a U1 gene from another lower eucaryote, *Physarum polycephalum* (26). The possibility that an essential transcription signal is present beyond the regions we have searched seems unlikely, since a fragment containing only 239 nt of 5′-flanking sequence and 120 nt of downstream DNA could restore viability to our U1 gene disruption.

The lack of significant sequence identity downstream of *S. pombe* snRNA genes suggests that initiation of transcription and 3′-end formation may not be linked in this organism, since this interdependence in mammals is mediated by a conserved box just 3′ to the snRNA coding sequence (9). A temperature-sensitive *S. pombe* mutant which accumulates 3′-extended snRNAs (33) may shed light on this question.

U1 snRNA is essential for viability. Each of the spliceosomal snRNAs from *S. cerevisiae*, including U1, has been shown to be essential for viability (1, 8, 29, 37, 39, 40), and recently U4 from *S. pombe* was likewise demonstrated to be indispensable for growth (11). It was thus not unexpected that disruption of the *S. pombe* U1 gene is lethal. Point mutations in U1 snRNA can now be assayed for their effects on snRNP assembly and splicing, which should help to clarify discrepancies between the *S. cerevisiae* and mammalian systems, and perhaps identify features which are unique to *S. pombe*.

ACKNOWLEDGMENTS

We thank David Beach (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) for *S. pombe* strains and plasmids. Our colleagues Claudia Reich and Henry Skinner are gratefully acknowledged for helpful discussions and comments on the manuscript.

This research was supported by Public Health Service grant RO1 GM38070 from the National Institutes of Health awarded to J.A.W.

LITERATURE CITED

1. **Ares, M.** 1986. U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. Cell **47**:49–59.
2. **Barke, D. G., J. H. M. White, and L. H. Johnston.** 1987. Molecular characterisation of the DNA ligase gene, *CDC17*, from the fission yeast *Schizosaccharomyces pombe*. Eur. J. Biochem. **162**:659–667.
3. **Booher, R., and D. Beach.** 1989. Involvement of a type 1 protein phosphatase encoded by *bws1*$^+$ in fission yeast mitotic control. Cell **57**:1009–1016.
4. **Branlant, C., A. Krol, and J.-P. Ebel.** 1981. The conformation of chicken, rat and human U1A RNAs in solution. Nucleic Acids Res. **9**:841–858.
5. **Branlant, C., A. Krol, J.-P. Ebel, E. Lazar, B. Haendler, and M. Jacob.** 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J. **1**:1259–1265.
6. **Brennwald, P., X. Liao, K. Holm, G. Porter, and J. A. Wise.** 1988. Identification of an essential *Schizosaccharomyces pombe* RNA homologous to the 7SL component of signal recognition particle. Mol. Cell. Biol. **8**:1580–1590.
7. **Brennwald, P., G. Porter, and J. A. Wise.** 1988. U2 small nuclear RNA is remarkably conserved between *Schizosaccharomyces pombe* and mammals. Mol. Cell. Biol. **8**:5575–5580.
8. **Brow, D., and C. Guthrie.** 1988. Spliceosomal snRNA U6 is remarkably conserved from yeast to mammals. Nature (London) **334**:213–218.
9. **Dahlberg, J. E., and E. Lund.** 1988. The genes and transcription of the major small nuclear RNAs, p. 38–70. *In* M. L. Birnstiel (ed.), Structure and function of major and minor snRNPs. Springer-Verlag KG, Berlin.
10. **Dandekar, T., and D. Tollervey.** 1989. Cloning of *Schizosaccharomyces pombe* genes encoding U1, U2, U3 and U4 snRNAs. Gene **81**:227–235.
11. **Dandekar, T., and D. Tollervey.** 1989. *Schizosaccharomyces pombe* U4 small nuclear RNA closely resembles vertebrate U4 and is required for growth. J. Mol. Biol. **208**:371–379.
12. **Guthrie, C., and B. Patterson.** 1988. Spliceosomal snRNAs. Annu. Rev. Genet. **22**:387–419.
13. **Gutz, H., H. Heslot, U. Leupold, and N. Loprieno.** 1974. *Schizosaccharomyces pombe*, p. 395–446. *In* R. C. King (ed.), Handbook of genetics. Plenum Publishing Corp., New York.
14. **Hamm, J., V. L. van Santen, R. A. Spritz, and I. W. Mattaj.** 1988. Loop I of U1 small nuclear RNA is the only essential RNA sequence for binding of specific U1 small nuclear ribonucleoprotein particle proteins. Mol. Cell. Biol. **8**:4787–4791.
15. **Kaufer, N. F., V. Simanis, and P. Nurse.** 1985. Fission yeast *Schizosaccharomyces pombe* correctly excises a mammalian RNA transcript intervening sequence. Nature (London) **318**: 78–80.
16. **Kiss, T., M. Antal, H. Hegyi, and F. Solymosy.** 1988. Plant small nuclear RNAs IV. The structure of U1 RNA from *Chlorella saccharophila*: a phylogenetic support, in terms of RNA structure, for the probable interaction between U1 and U2 snRNPs during the splicing of pre-mRNA. Nucleic Acids Res. **16**:2734.
17. **Kretzner, L., A. Krol, and M. Rosbash.** 1989. *S. cerevisiae* U1 snRNA secondary structure contains both universal and yeast-specific domains. Proc. Natl. Acad. Sci. USA **87**:851–855.
18. **Kretzner, L., B. C. Rymond, and M. Rosbash.** 1987. *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. Cell **50**:593–602.
19. **Liao, X., P. Brennwald, and J. A. Wise.** 1989. Genetic analysis of *Schizosaccharomyces pombe* 7SL RNA: a structural motif that includes a conserved tetranucleotide loop is important for function. Proc. Natl. Acad. Sci. USA **86**:4137–4141.
20. **Losson, R., and F. Lacroute.** 1983. Plasmids carrying the yeast OMP decarboxylase structural and regulatory genes: transcription regulation in a foreign environment. Cell **32**:371–377.
21. **Lührmann, R.** 1988. snRNP proteins, p. 71–99. *In* M. L. Birnstiel (ed.), Structure and function of major and minor snRNPs. Springer-Verlag KG, Berlin.
22. **Lutz-Reyermuth, C., and J. D. Keene.** 1989. The U1 RNA-binding site of the U1 small nuclear ribonucleoprotein (snRNP)-associated A protein suggests a similarity with U2 snRNPs. Mol. Cell. Biol. **9**:2975–2982.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Mattaj, I. W., and E. M. DeRobertis.** 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell **40**:111–118.
25. **Mitchison, J. M.** 1970. Physiological and cytological methods of *Schizosaccharomyces pombe*. Methods Cell Physiol. **4**:131–156.
26. **Myslinski, E., F. X. Wilhelm, and C. Branlant.** 1989. A structural analysis of *P. polycephalum* U1 RNA at the RNA and gene levels. Are there differentially expressed U1 RNA genes in *P. polycephalum*? U1 RNA evolution. Nucleic Acids Res. **17**: 1019–1034.
27. **Ohkura, H., K. Kinoshita, S. Miyatani, T. Toda, and M. Yanagida.** 1989. The fission yeast *dis2*$^+$ gene required for chromosome disjoining encodes one of two putative type 1 protein phosphatases. Cell **57**:997–1007.
28. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
29. **Patterson, B., and C. Guthrie.** 1987. An essential yeast snRNA with a U5-like domain is required for splicing *in vivo*. Cell **49**:613–624.
30. **Patton, J. R., W. Habets, W. J. van Venrooij, and T. Pederson.** 1989. U1 small nuclear ribonucleoprotein particle-specific proteins interact with the first and second stem-loops of U1 RNA, with the A protein binding directly to the RNA independently of the 70K and Sm proteins. Mol. Cell. Biol. **9**:3360–3368.
31. **Patton, J. R., and T. Pederson.** 1988. The M_r 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5′ stem-loop of U1 RNA and interacts with Sm domain proteins. Proc. Natl. Acad. Sci. USA **85**:747–751.
32. **Porter, G., P. Brennwald, K. Holm, and J. A. Wise.** 1988. The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes. Nucleic Acids Res. **16**:10131–10152.
33. **Potashkin, J., and D. Frendewey.** 1989. A mutation in a single gene of *Schizosaccharomyces pombe* affects the expression of several snRNAs and causes defects in RNA processing. EMBO J. **9**:525–534.
34. **Query, C. C., R. C. Bentley, and J. D. Keene.** 1989. A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. Mol. Cell. Biol. **9**:4872–4881.
35. **Russell, P.** 1983. Evolutionary divergence of mRNA transcription initiation mechanism in yeast. Nature (London) **301**:167–169.
36. **Russell, P., and P. Nurse.** 1986. *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*: a look at yeasts divided. Cell **45**:781–782.
37. **Seraphin, B., 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.
38. **Seraphin, B., and M. Rosbash.** 1989. Mutational analysis of the interactions between U1 small nuclear RNA and pre-mRNA of yeast. Gene **82**:145–151.
39. **Siliciano, P. G., D. A. Brow, H. Roiha, and C. Guthrie.** 1987. An essential snRNA from *S. cerevisiae* has properties predicted for U4, including interaction with a U6-like snRNA. Cell **50**: 585–592.
40. **Siliciano, P. G., and C. Guthrie.** 1988. 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal addi-

tional requirements. Genes Dev. **2**:1258–1267.
41. **Siliciano, P. G., M. H. Jones, and C. Guthrie.** 1987. *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. Science **237**:1484–1487.
42. **Sillekens, P. T. G., W. J. Habets, R. P. Beijer, and W. J. van Venrooij.** 1987. cDNA cloning of the human U1 snRNP-specific A-protein: extensive homology between U1 and U2 snRNP-specific proteins. EMBO J. **6**:3841–3848.
43. **Small, K., P. Brennwald, H. Skinner, K. Schaefer, and J. A. Wise.** 1989. Sequence and structure of U5 snRNA from *Schizosaccharomyces pombe*. Nucleic Acids Res. **17**:9483.
44. **Steitz, J. A., D. L. Black, V. Gerke, K. Parker, A. Kramer, D. Frendewey, and W. Keller.** 1988. Functions of the abundant U-snRNPs, p. 115–154. *In* M. L. Birnstiel (ed.), Structure and function of major and minor snRNPs. Springer-Verlag KG, Berlin.
45. **Surowy, C. S., V. L. van Santen, S. M. Scheib-Wixted, and R. A. Spritz.** 1989. Direct, sequence-specific binding of the human U1-70K ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. Mol. Cell. Biol. **9**:4179–4186.
46. **Tani, T., and Y. Ohshima.** 1989. The gene for U6 small nuclear RNA in fission yeast has an intron. Nature (London) **337**:87–90.
47. **Vankan, P., D. Edoh, and W. Filipowicz.** 1988. Structure and expression of the U5 snRNA gene of *Arabidopsis thaliana*. Conserved upstream sequence elements in plant U-RNA genes. Nucleic Acids Res. **16**:10425–10440.
48. **Wu, H. N., and O. C. Uhlenbeck.** 1987. Role of a bulged A residue in a specific RNA-protein interaction. Biochemistry **26**:8221–8227.
49. **Yuo, C.-Y., and A. M. Weiner.** 1989. Genetic analysis of the role of human U1 snRNA in mRNA splicing: I. Effect of mutations in the highly conserved stem-loop I of U1. Genes Dev. **3**: 697–707.
50. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46**:824–835.