

U1 Precursors: Variant 3' Flanking Sequences Are Transcribed in Human Cells

James G. Patton and Eric D. Wieben

Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, Minnesota 55905

Abstract. Using RNase protection and oligonucleotide hybridization experiments, we have shown that U1 precursors are derived by transcription of 3' flanking sequences. A labeled SP6 transcript of one of the true U1 genes (pD2) was able to protect a subset of the 3' flanking sequences present in HeLa cytoplasmic U1 RNA. However, not all U1 precursors were protected using this probe, suggesting that variant U1 precursor 3' tail sequences are expressed in HeLa cells. This

conclusion has been confirmed by hybridization of HeLa RNA samples with specific oligonucleotide probes representing variant U1 3' flanking sequences. Interestingly, these variant tail sequences contain the putative Sm antigen binding site, A(U)₃₋₆G. The conservation of this flanking sequence through evolution suggests a possible functional role for these precursor tails in ordering protein binding to U1 RNA.

U1 RNA is a highly abundant, small nuclear RNA found in the nuclei of eucaryotic cells. It is complexed with 8–10 proteins (14, 16, 47) forming a ribonucleoprotein (RNP)¹ particle (snRNP). The finding that U1 RNA contains sequences complementary to intron–exon splice junctions in pre-messenger RNAs led to the hypothesis that U1 RNA might be required for mRNA splicing (22, 41). This hypothesis was supported by the finding that antibodies directed against U1 snRNP inhibit both in vitro and in vivo splicing (18, 39, 51). Further confirmation that U1 is required for splicing comes from experiments demonstrating that specific degradation of the 5' end of U1 RNA also inhibits in vitro splicing (4, 17, 18). The fact that antibodies to U1 snRNP recognize the protein components of U1 RNP and not U1 RNA alone, along with the finding that only U1 RNP will bind specifically to 5' splice sites (36) suggests that it is the RNP form of U1 that functions in splicing.

In the human genome there are estimated to be ~1,000–2,000 copies of U1 RNA genes (9, 31). Of these, 30–125 are thought to represent true U1 genes, while the remaining copies are thought to be pseudogenes (25, 31). The set of true U1 genes are denoted as true genes based on sequence identity with a human U1 RNA sequenced by Branlant et al., (5) and the ability to be transcribed both in *Xenopus* oocytes and in vitro (25, 31).

Like other RNAs, U1 is apparently synthesized as a precursor (11, 29, 52). Recent experiments suggest that U1 precursors are transported to the cytoplasm, where they associate with snRNP proteins and undergo 3' terminal processing before returning to the nucleus. Since no introns have been found in U1 genes, and transcription apparently begins at the

cap site (43), it has been presumed that the precursor-specific sequences are derived by transcription of 3' flanking sequences. In further support of this hypothesis, recent experiments have established that some read-through transcription beyond the 3' end of the sequence found in mature U1 does occur (20), and sequences responsible for the formation of 3' ends have been identified (13).

We have been using RNase mapping to investigate the structure of U1 precursors in greater detail. Since all of the true genes that have been sequenced have identical 3' flanking sequences for the first 19 nucleotides (nt) after the final nucleotide found in mature U1 RNA, it was surprising to find that a cloned copy of one of the true U1 genes (pD2; reference 31) does not protect the total pattern of cytoplasmic U1 precursors from RNase digestion when hybridized to these RNAs. This suggests that not all U1 precursors are derived via transcription of the true U1 genes. Hybridization of HeLa cell U1 RNA to specific oligonucleotides containing variant 3' flanking sequences confirms this result, and suggests that the regulation of U1 RNA synthesis in human cells might be much more complex than was previously supposed.

Materials and Methods

Cell Lines, RNA Isolation, and Antisera

HeLa cells were grown in suspension culture in Joklik's modified minimum essential media (Gibco, Grand Island, NY) supplemented with 5% fetal calf serum. Cells were pulse-labeled for 45 min with [³H]uridine as described (29) or labeled for 12–16 h with [³H]uridine for long labeling studies and were then fractionated in RSB (0.01 M NaCl, 0.01 M Tris-HCl, pH 7.2, and 1.5 mM MgCl₂) as described (40). We have experienced some variability in the purity of cell fractions obtained by this procedure. Given this variability, plus the possibility of nuclear leakage during fractionation, we wish to emphasize that the terms nuclear and cytoplasmic reflect operational definitions, and do not necessarily reflect the true cellular localization of particu-

1. *Abbreviations used in this paper:* RNP, ribonucleoprotein; snRNP, small nuclear ribonucleoprotein particle.

lar molecules. Total RNA was isolated from whole cells or isolated fractions by phenol-chloroform extraction. Where indicated, U1 RNA was isolated by antibody precipitation using protein A-Sepharose chromatography followed by phenol extraction as described (45-47). Hybrid selection of U1 RNA was as described (29). Antisera used in this study were obtained from patients with systemic lupus erythematosus. IgG was isolated from these patients and was found to specifically precipitate U1 snRNPs.

Oligonucleotide Synthesis

Oligodeoxynucleotides were synthesized using the phosphoramidite method (2) on a synthesizer (380A; Applied Biosystems, Inc., Foster City, CA). Controlled pore glass served as the solid support (1).

SP6 Transcriptions

A human U1 gene, pD2, (31) was cloned into plasmids pSP64 and pSP65 (Promega Biotech, Madison, WI). Plasmid SP64U1 contains the Bgl II-Eco RI fragment of pD2 cloned into the Bam HI and Eco RI sites in pSP64. Plasmid pSP65cU1 contains the Bgl II-Rsa I fragment of pD2 cloned into the Bam HI and Sma I sites in pSP65. RNA transcripts were generated from these plasmid constructs using SP6 polymerase (Boehringer Mannheim Diagnostics, Inc., Houston, TX) following the manufacturers instructions (Promega Biotech). The transcript from pSP64U1 contains U1 RNA beginning at the Bgl II site at -6 in addition to 34 nucleotides of polylinker sequences at the 5' end. An RNA transcript complementary to U1 was generated from pSP65cU1 with transcription of 25 polylinker sequences followed by cU1 RNA beginning at the Rsa I site (+194). Transcription was terminated at convenient restriction sites as described in the figure legends.

RNase Digestion Experiments

SP6-generated transcripts were hybridized to RNA isolated from HeLa cells in a buffer containing 0.4 M NaCl, 40 mM Pipes, pH 6.4, 1.0 mM EDTA, and 80% Formamide. Samples were briefly denatured at 80°C and hybridized for 12-16 h at 50°C. Unhybridized sequences were then digested for 1 h at 24°C in RNase digestion buffer (10 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.3 M NaCl, 2.0 µg/ml RNase T1, and 40 µg/ml RNase A) followed by treatment with 0.1 µg/ml proteinase K and SDS (0.6%) for 15 min at 37°C. Remaining undigested RNAs were phenol-chloroform extracted in the presence of carrier tRNA (5 µg) and ethanol precipitated. Reconstituted samples were run on a 10% polyacrylamide gel containing 7.0 M urea. Gels were electrophoresed for 2-3 h at 20 mA, and either exposed directly or fluorographed as described (46, 47).

Oligonucleotide Hybridizations

RNAs were incubated in 3 vol of 6.15 M formaldehyde, 10× SSC (1× is 150 mM NaCl, 15 mM Sodium Citrate) at 65°C for 10 min and subsequently

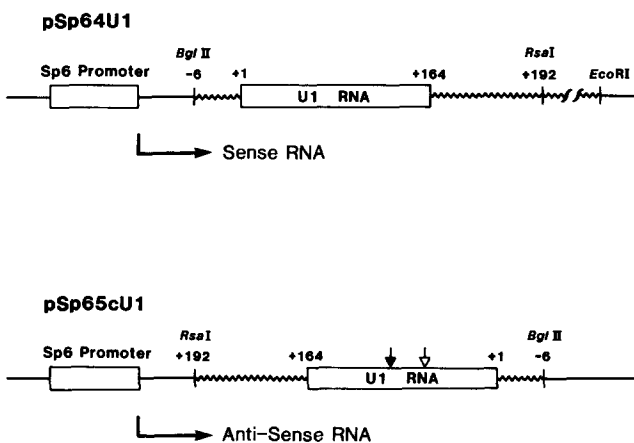


Figure 1. Construction of plasmids pSP64U1 and pSP65cU1. A human U1 RNA gene (pD2, reference 31) was subcloned into plasmids pSP64 and pSP65 as detailed in Materials and Methods. The solid arrow refers to the Taq I site at position +115 and the open arrow refers to the Hpa II site at +73.

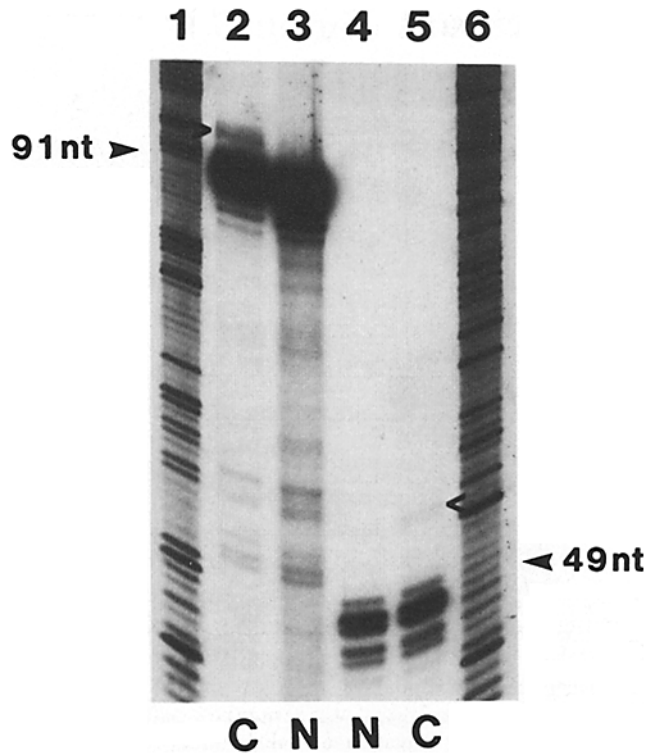


Figure 2. The 3' flanking sequence of the human U1 RNA true gene, pD2, is expressed in HeLa cells. U1 RNAs were isolated from nuclear (lanes 3 and 4) and cytoplasmic (lanes 2 and 5) fractions of HeLa cells using anti(U1)RNP antisera followed by phenol extraction. These RNAs were hybridized with ³²P-labeled RNA probes synthesized from an SP6 vector (see Fig. 1) containing the true human U1 gene, pD2, which had been digested with either Hpa II (lanes 2 and 3) or Taq I (lanes 4 and 5). After digestion of unhybridized sequences with RNases A and T1, protected RNAs were purified by phenol extraction and ethanol precipitation. Lanes 1 and 6 contain markers derived from DNA sequencing reactions. Samples were electrophoresed on a 10% acrylamide/7.0 M urea gel. The arrowheads denote the precursor-specific protected bands. nt, nucleotide.

slot-blotted onto nitrocellulose (Schleicher and Schuell, Inc., Keene, NH). Filters were baked at 80°C for 2 h. Prehybridization was for 4 h at 45°C in 0.9 M NaCl, 0.18 M Tris-HCl, pH 8.0, 12 mM EDTA, 1% nonfat dry milk, 1% SDS, and 250 µg/ml tRNA. Hybridizations were in the same buffer containing 2 × 10⁶ cpm/ml for 12-24 h at 45°C. Blots were washed at room temperature in 6× SSC for 1-2 h followed by two washes for 20 min each at 37°C in 3.0 M tetramethylammonium chloride, 50 mM Tris-HCl, pH 8.0, 2.0 mM EDTA, and 0.1% SDS. The final wash solution and temperature were determined by following the procedure of Wood et al. (50). Oligonucleotides were end labeled using polynucleotide kinase (Boehringer Mannheim Diagnostics, Inc.) and [γ-³²P]ATP as described (30).

Results

The 3' Flanking Sequence of the Human U1 RNA True Gene, pD2, Is Expressed in HeLa Cells

Precursors to U1 RNA that are 3-15 nucleotides longer than mature U1 (11, 29, 52) have been described. To determine definitively whether these precursors are derived via transcription of 3' flanking sequences, we performed RNase digestion experiments. Using the SP6 transcription system, we prepared ³²P-labeled RNA transcripts complementary to one of the true U1 RNA genes (pD2, reference 31). As seen

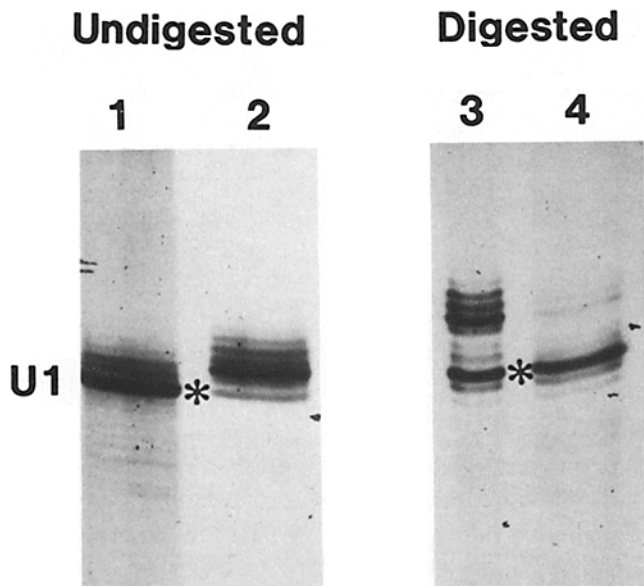


Figure 3. Only a subset of the U1 precursors are transcribed from a pD2-homologous U1 gene. U1 RNA was isolated from nuclear (lane 1) and cytoplasmic (lane 2) fractions of pulse-labeled HeLa cells using anti(U1)RNP antisera. These RNAs were hybridized to an unlabeled SP6 transcript that was terminated at the Hpa II site in pSP64cU1. After digestion of unhybridized sequences with RNases A and T1, protected RNAs were purified and electrophoresed as in Fig. 1 except that the gel was fluorographed and dried before exposing to film. Lanes 1 and 2, undigested RNAs isolated from nuclear (lane 1) and cytoplasmic (lane 2) fractions. Lanes 3 and 4, protected RNAs from cytoplasmic (lane 3) and nuclear (lane 4) fractions. All samples were run on the same gel. The two pairs of lanes have been aligned in the figure to permit comparison of the patterns of labeled RNAs. The asterisks denote the position of mature U1 (lanes 1 and 2) or protected species derived from mature U1 (lanes 3 and 4).

in Fig. 1, RNA transcripts were prepared that terminated at either the Taq I site (+115) or the Hpa II site (+73). These transcripts contain 28 bases of 3' flanking sequences as well as 49 bases (Taq I probe) or 91 bases (Hpa II probe) of mature U1 RNA complementary sequences.

U1 RNAs were isolated from nuclear or cytoplasmic fractions of HeLa cells and hybridized to the SP6 transcripts, followed by RNase digestion with RNases A and T1. As seen in Fig. 2, both cytoplasmic and nuclear U1 RNA samples yield protected RNA fragments that represent hybridization to mature U1 RNA sequences (48–49 nucleotides with the Taq I probe and 90–91 nucleotides with the Hpa II probe). Protected species 6–7 nucleotides longer than expected for mature U1 sequences are found when cytoplasmic, but not nuclear, RNAs are hybridized with either of the pD2-derived probes. Therefore, these HeLa cytoplasmic fractions contain U1 RNA species that are extended at their 3' terminus by up to seven nucleotides. Furthermore, these species must have been derived by transcription of a pD2-homologous 3' flanking sequence.

Not all U1 RNA Precursors Are Transcribed from a pD2-homologous U1 Gene

The finding that 3' flanking sequences from prototypic U1 genes are expressed in HeLa cells agrees with the results of

a previous S1 mapping experiment (13), but the size distribution of the protected species is somewhat different than the pattern of U1 obtained by pulse-labeling (Fig. 3, lanes 1 and 2). However, the RNase mapping experiment shown in Fig. 2 reveals only the most stable U1 species present in the cytoplasmic fraction. Because of the difference in methodology, it is difficult to determine exactly which of the precursor RNAs are synthesized from pD2-homologous genes. To address this problem, an unlabeled SP6 probe was synthesized from Hpa II-digested pSP65cU1. This was hybridized to pulse-labeled, antibody-purified U1 RNAs, and the unhybridized sequences were digested with RNases A plus T1 as before. By restricting the analysis to the pulse-labeled RNAs, it is possible to directly compare the pattern of protected RNAs with the pattern of U1 precursors identified by pulse-chase analysis, antibody selection, and hybridization analysis.

Lanes 1 and 2 of Fig. 3 illustrate the pattern of pulse-labeled RNAs recovered from HeLa cells with anti(U1)RNP antibody. The pattern of protected RNAs is shown in lanes 3 and 4 of Fig. 3. If all of the U1 precursors present in lanes 1 and 2 are derived by transcription of 3' flanking sequences of pD2-homologous genes, then the pattern of protected species should be identical to the total pattern of U1 precursors used for the RNase protection experiment. The results shown in Fig. 3 demonstrate that the pD2 gene protects the mature U1 portion of the precursors, but only a subset of 3' tails from digestion by RNase A plus T1. Thus, there is a large accumulation of radioactive species in lane 3 that migrate at the position of protected mature U1 sequences. The fact that there is a specific increase in the material that migrates at the position of protected mature U1 sequences suggests that the shift in pattern is not due to artifactual nibbling of the nuclease at perfectly matched hybrids. This conclusion is supported by the work of other groups, who have found that digestion with RNase A and T1 can be used to detect single-base mismatches in both RNA–DNA and RNA–RNA duplexes with a minimum of nonspecific activity (34, 37, 49). However, since neither nuclease will cleave after mismatched A residues, it is difficult to draw further conclusions from the pattern of protected species produced by this type of analysis. In the case of the experiment shown in Fig. 3, this difficulty is compounded by the fact that the RNA samples used in this experiment were labeled with [³H]uridine for only 25 min in vivo. Hence, the patterns of labeled RNA shown in this figure do not reflect the steady-state populations of U1 precursors from HeLa cells. Nevertheless, it is clear that the SP6 true gene probe does not protect the entire ladder of pulse-labeled U1 precursors from nuclease digestion. This result suggests that some 3' tails on U1 precursors are not derived by transcription of the 3' flanking sequence found in pD2. Further experiments using specific oligonucleotide probes confirm this conclusion.

Variant 3' Flanking Sequences Are Transcribed in HeLa Cells

If only some of the U1 precursors have 3' tails corresponding to the 3' flanking sequences found in pD2, what are the sequences of the remaining species? Examination of the published sequences of five class I U1 pseudogenes reveals that despite extensive homology within both coding and flanking regions, the immediate 3' flanking sequences differ from

U1 Gene Sequence		Oligomer	
5'-TTTCCCCTGACTTTCTG-3'	pD2	3'-AAAGGGGACTGAAAGAC-5'	Plusmer *
5'-TTTCCCCTGA <u>A</u> TTTTITG-3'	pU1-6	3'-AAAGGGGACTTAAAAAC-5'	IO **
5'-TTTCCCCTGA <u>I</u> TTTTITG-3'	pU1.1	3'-AAAGGGGACTTAAAAAC-5'	CO ***
5'-TTTCCCCTGA <u>I</u> TTTTITG-3'	pU1.15		***

Figure 4. Oligonucleotides complementary to variant U1 3' flanking sequences. The underlined nucleotides designate the positions of divergence from the prototypic U1 sequence. *Reference 31; **reference 35; ***reference 9.

those found in pD2-homologous true genes in every case (9, 35). Two particular 3' flanking sequence variants are found in multiple pseudogenes (Fig. 4). Instead of the prototypic flanking sequence -ACTTTCTG, two pseudogenes (U1.1 and U1.15) have the sequence -ATTTTTTG, and two more (U1.4 and pU1-6) have a sequence of the form -AATTTT(T)G. The occurrence of these variant sequences in more than one pseudogene sequence raises the possibility that functional U1 genes exist that have these flanking sequences. To directly test this hypothesis, we synthesized oligodeoxynucleotide probes complementary to these variant sequences as detailed in Fig. 4.

Since our data suggest that the most prevalent precursors contain only four to eight extra nucleotides at the 3' end of the mature U1 sequence, the probes were designed to overlap the mature sequence for nine nucleotides. This overlap is not in itself sufficient to allow hybridization to mature U1 sequences under the conditions used, but it does provide a long enough contiguous stretch of sequence to prime hybridization within the immediate 3' flanking sequences. The oligonucleotides contain mismatches with the pD2 gene at the underlined positions. These mismatches are sufficient to prevent hybridization of the oligonucleotides to an SP6 tran-

script of the pD2 gene sequence (see Materials and Methods for hybridization conditions). The specificity of hybridization is shown in Fig. 5. Total RNA from nuclear and cytoplasmic fractions of HeLa and K562 erythro leukemia cells was slot-blotted alongside an SP6 transcript derived from pD2 (see Fig. 1) and hybridized to the end-labeled oligonucleotides. As shown, the only oligonucleotide that hybridizes to the SP6 transcript is 3'-AAAGGGGACTGAAAGAC-5' (Plusmer), which is completely complementary to pD2. None of the probes hybridize to tRNA, confirming the specificity of the hybridization and washing conditions. Thus, hybridization under these conditions requires more than the 10 contiguous bases that are common to all three oligonucleotides. However, both 3'-AAAGGGGACTTAAAAAC-5' (IO) and 3'-AAAGGGGACTAAAAAC-5' (CO) (see Fig. 4) hybridize to both HeLa and K562 total RNA. This result suggests that sequences complementary to these oligonucleotides are expressed in both of these two cell lines.

In the samples shown, there is significantly more hybridizing material in the HeLa nuclear samples than in the corresponding samples from K562 cells. We have noted some variability in the relative intensity of hybridizing species in

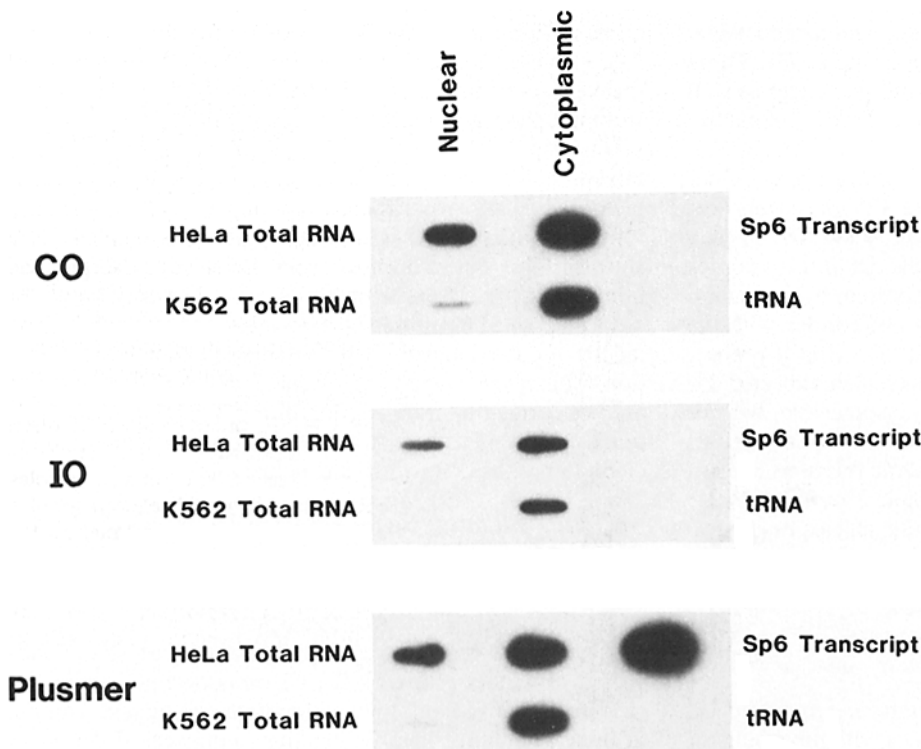


Figure 5. Hybridization of end-labeled oligonucleotides to HeLa total RNA and K562 total RNA. Total RNA was isolated from nuclear and cytoplasmic fractions of HeLa cells and K562 erythro leukemia cells by phenol extraction. 10 µg of nuclear RNA and 40 µg of cytoplasmic RNA were loaded onto slot blots along with an aliquot of a U1 RNA-homologous SP6 transcript from Rsa I-digested pSP64U1. 10 µg of tRNA was also blotted as a background control. Oligonucleotides were end labeled and hybridized to total RNA as described in Materials and Methods. All of the samples are from the same hybridizations and represent identical exposures (2 d).

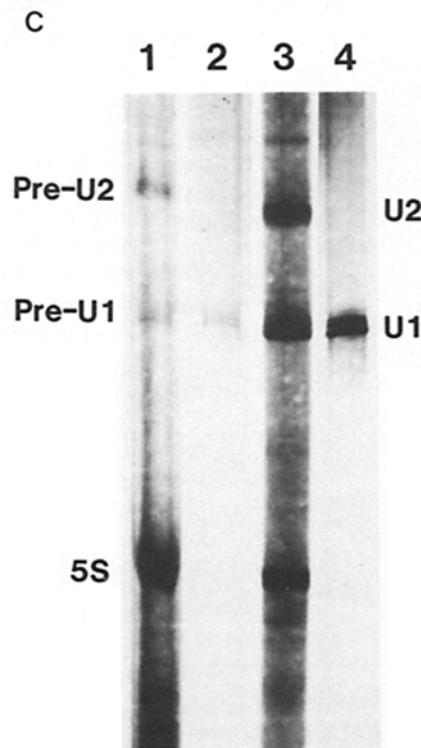
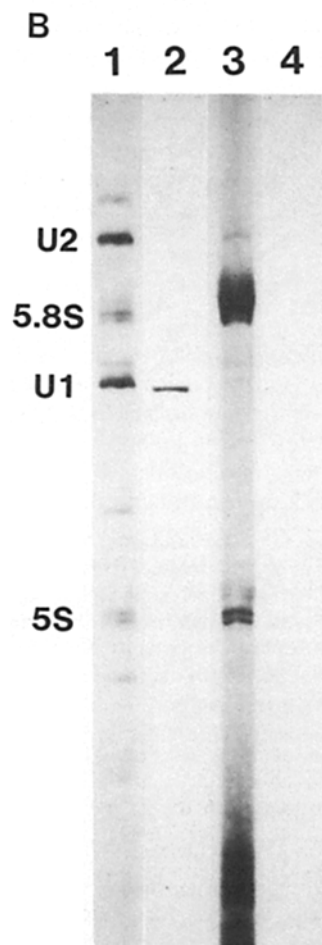
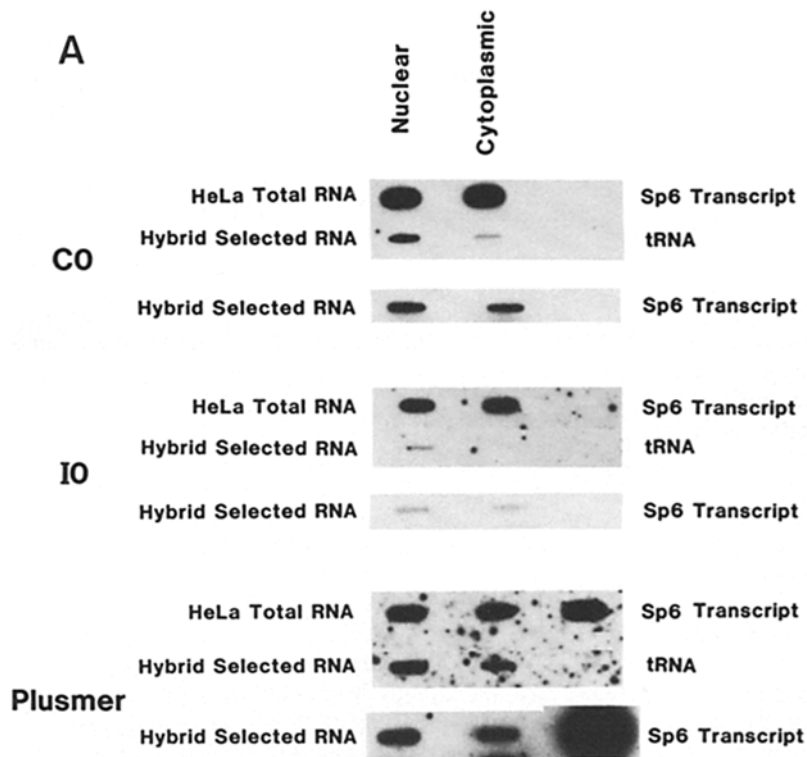


Figure 6. IO and CO recognize U1-specific sequences other than the true U1 RNA sequence. Total RNA was isolated from pulse-labeled and overnight-labeled (**B**) HeLa cells. These RNAs were hybridized to nitrocellulose-immobilized M13 DNA containing sequences complementary to U1 RNA. Hybrid-selected sequences were eluted from the filters, phenol extracted, and ethanol precipitated. Equal aliquots of the reconstituted RNAs were then slot blotted (**A**) or run on a 10% acrylamide gel and fluorographed (**B** and **C**). (**A**) 10 μ g of unlabeled total nuclear RNA, 40 μ g of unlabeled total cytoplasmic RNA, 10 μ g of tRNA, equal aliquots of SP6 transcripts, and an equal amount of hybrid-selected RNA loaded onto the gels in **B** and **C** were slot blotted onto nitrocellulose. Hybridization with end-labeled oligonucleotides and conditions were as described in Materials and Methods. The bottom row of hybrid-selected samples contains the pulse-labeled material shown in **C**. These samples were hybridized separately from the samples shown in the top two rows of each group. Electrophoretic analysis of the nonselected RNAs (not shown) indicated that the selections did not quantitatively remove U1 RNAs from the total RNA samples. Thus, quantitative comparisons between the three groups of samples (total RNA, long-labeled selections, and short-labeled selections) are not meaningful. (**B**) Fluorograph of RNA isolated from HeLa cells labeled with [3 H]uridine for 16 h. Lane 1, 6 μ g of total nuclear RNA; lane 2, hybrid-selected nuclear RNA; lane 3, 10 μ g of total cytoplasmic RNA; lane 4, hybrid-selected cytoplasmic RNA. The exposure time in each case was 30 d. (**C**) Fluorograph of RNA isolated from HeLa cells labeled with [3 H]uridine for 45 min. Lane 1, 10 μ g of total cytoplasmic RNA; lane 2, hybrid-selected cytoplasmic RNA; lane 3, 10 μ g of total nuclear RNA; lane 4, hybrid-selected nuclear RNA. Lanes 1 and 3 are from an 18-h exposure of the autoradiogram. Lanes 2 and 4 are from a 30-d exposure of the same autoradiogram.

cytoplasmic and nuclear fractions from HeLa cells (compare to Fig. 6 A), which we suspect is due to differences in the efficiency of cell fractionation. Accordingly, we wish to emphasize the operational nature of our designations of cellular localization. Given this limitation, we do not ascribe any significance to the differences noted between the two cell lines. However, the relative intensities of the signals obtained with the different oligonucleotides was quite reproducible. Although there was some variation in the specific activity of the three probes from experiment to experiment, hybridization with Plusmer and CO routinely yielded stronger signals than hybridization with IO. This not only provides some information regarding the relative abundance of these sequences in RNA from the two cell types, but provides confirmation that IO and CO are recognizing different sequences.

IO and CO Specifically Recognize U1 RNA Sequences

To determine if the hybridization of the variant oligonucleotides to total RNA was actually due to hybridization to U1 RNA sequences, U1 RNA was purified by hybridization selection before hybridization. Fig. 6 A shows that the hybrid-selected, pure U1 RNAs, hybridize to all three of the end-labeled oligonucleotides under conditions that do not permit hybridization of the variant oligonucleotides to the pD2 flanking sequences. Again, only Plusmer hybridizes to the SP6 transcript as well as to the selected RNA. Rehybridization of these filters to a coding region probe confirms the presence of equal quantities of the SP6 true U1 transcript on all three sets of filters (data not shown).

Since samples prepared by hybridization selection may contain contaminating RNA, we analyzed the purity of our hybrid-selected samples by electrophoresis. The RNA used for the hybrid selections shown in Fig. 6 A was isolated from HeLa cells that were prelabeled with [³H]uridine. Aliquots of the [³H]uridine-labeled, hybrid-selected samples equivalent to that used for the slot blots shown in Fig. 6 A were electrophoresed on 10% acrylamide gels and visualized by fluorography. As shown in Fig. 6, B and C, these selection experiments resulted in the isolation of pure U1 RNA from both nuclear and cytoplasmic samples. The long-labeled U1 produced a single band in the nuclear lane after fluorography, while the cytoplasmic precursor species were not visible (Fig. 6 B). (The low recovery of material from the cytoplasmic sample in this experiment is consistent with the low level of hybridization observed in this sample in Fig. 6 A, top row of hybrid-selected samples.) To rule out the possibility that the hybrid-selected U1 samples were contaminated with unstable RNAs that would not be seen in long-labeled RNA samples, the hybrid selections were repeated with pulse-labeled RNAs (bottom row of hybrid-selected samples in Fig. 6, A and C). Although some diffuse radioactivity is present at the top of the hybrid-selected lanes in Fig. 6 C, the major species in these samples are clearly the normal pattern of U1 precursors. Taken as a whole, these results suggest that the variant oligonucleotides are specifically recognizing U1 RNA sequences.

The recovery of diffuse, higher molecular weight species in the pulse-labeled hybrid selection experiments raises the possibility that the variant oligonucleotides are hybridizing to unstable aberrant transcripts that may not be physiologically active. To evaluate this possibility, and to provide fur-

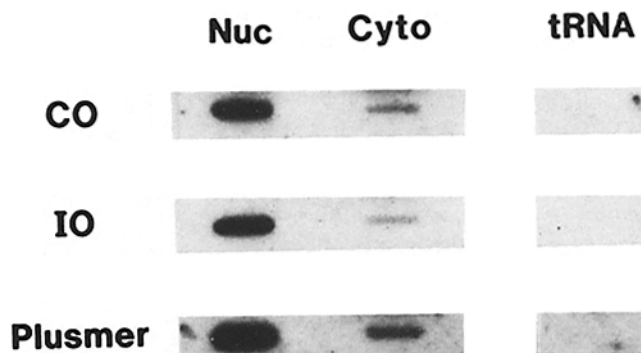


Figure 7. IO and CO recognize RNP antisera-selected RNA. RNA was isolated from unlabeled HeLa cells using U1-specific antisera (RNP). Equal aliquots of antibody-selected RNAs were slot blotted onto nitrocellulose along with 10 μ g of tRNA and an equal amount of an SP6 transcript of U1 RNA. Hybridization was as in Fig. 6. The SP6 blot can be seen in the bottom row of each group in Fig. 6 A. Exposure time was identical to that shown in Fig. 6 A.

ther confirmation of the specificity of the hybridization conditions, we also analyzed the hybridization of the variant oligonucleotides to U1 RNAs isolated with anti(U1)RNP antisera. This strategy confines the analysis to U1 molecules that have been assembled into ribonucleoprotein complexes. In agreement with the results outlined above, antibody-purified U1 RNA from both nuclear and cytoplasmic cell fractions does hybridize to the variant sequences under conditions that do not permit hybridization of the variants to the pD2 flanking sequences (Fig. 7). (The SP6 controls for these samples are shown in Fig. 6 A). This result is consistent with the hypothesis that the variant U1 sequences are packaged into functional RNP complexes.

Discussion

Our investigation of the sequence structure of U1 precursors has revealed that multiple precursor species up to seven nucleotides longer than mature U1 are transcribed from genes with 3' flanking sequences identical to those found in the characterized true U1 genes. This finding is in good agreement with previous studies that have found that transcription of prototypic U1 genes begins at the cap site (43) and extends past the end of the mature U1 sequence (20). The current work also provides an important link between these studies of U1 transcription and the previous description of multiple, unstable U1 species found in cytoplasmic fractions from HeLa cells (29). The RNase mapping studies demonstrate that the multiple species represent true length heterogeneity of the pre-U1 species, and that these U1 precursors are indeed longer than the +3 nucleotide species described by others (11, 52).

An unexpected result of the RNase mapping experiments was that not all of the precursor species identified by pulse labeling and antibody precipitation were direct transcriptional extensions from prototypic U1 true genes. This finding was confirmed by hybridization of U1 RNA samples to two oligonucleotides complementary to specific variants of 3' flanking sequences found in previously characterized pseudogenes. These experiments indicate that U1 genes with vari-

ant 3' flanking sequences are transcribed in HeLa cells and that transcripts containing these variant flanking sequences become packaged with proteins to form ribonucleoprotein complexes. We wish to emphasize that there may be significantly more heterogeneity in the 3' tails of U1 precursors than we have identified with the two oligonucleotide probes used in this study.

One of the important questions that our studies have not addressed is whether genes previously characterized as class I pseudogenes can be transcribed *in vivo* to produce functional transcripts. Although the sequences used to construct probes were taken from published pseudogene sequences, it is possible that there are functional U1 genes that have variant 3' flanking sequences while coding for a mature RNA that is identical to the published sequence of U1 from HeLa cells (5). The alternative possibility, that some presumed class I pseudogenes are really transcribed in some cells, is equally interesting.

Variant U1 RNAs have been described in chickens (42), *Xenopus* (8, 12, 19), and mice (15, 27). However, since human U1 RNA from HeLa cells yielded an unambiguous sequence, genes that differ from the RNA sequence by as little as one nucleotide have been designated as pseudogenes. Nevertheless, neither the RNA sequencing data nor our RNase mapping studies exclude the possibility that there is scattered low level sequence heterogeneity in mature human U1 RNA. The regulated expression of U1 RNAs with variant sequences would have obvious implications for the regulation of alternative splicing events (27). Our experiments suggest that variant U1 flanking sequences are expressed in at least two different human tumor cell lines. We are investigating the possibility that these variants might be differentially regulated in different cell types.

In theory, it is not necessary to invoke heterogeneity within the coding sequence to produce a heterogeneous population of U1 RNPs. The formation of U1 ribonucleoprotein complexes *in vivo* involves the interaction of snRNP proteins with U1 precursors rather than mature U1 (29). Although it is not known whether complete assembly of U1 RNP requires the precursor-specific 3' tail, assembly experiments using mature U1 RNA have resulted in the formation of incomplete RNA-protein complexes (46). While there are many possible explanations for this result, one interesting possibility is that the 3' tail is transiently used *in vivo* as a protein binding site. According to this scenario, proteins bind to the 3' tail and order subsequent protein assembly within the mature region before the tail is lost upon processing. This hypothesis would predict that molecules with different 3' tails might become associated with different patterns of proteins. The observation that only a subset of pre-U1 snRNPs are recognized by patient anti-La antibodies (28) is entirely consistent with this hypothesis, as is the heterogeneity noted when intact snRNP particles are isolated (14, 16). The diversity in U1 precursors described here provides a possible basis for the decision as to which precursors acquire a given set of proteins.

It is particularly interesting in this regard that both the variant 3' flanking sequences from which we constructed our oligonucleotide probes contain the putative recognition site for the binding of the Sm antigen. The sequence A(U)_nG with *n* ranging from 3–6 is required for the binding of the Sm antigenic proteins to U2 RNA in *Xenopus* (33), and a

Table I. Evolutionary Conservation of Sm Binding Sites at the 3' Ends of U1 Genes

Species	Locus	3' Sequence	Reference
Human	HU1-1	–ACTTCTG	(31)
Human	U1.4	–AATTTTG	(9)
Human	U1.1	–ATTTTTTG	(9)
Human	U1.15	–ATTTTTTG	(9)
Human	pU1-6	–AATTTTTG	(35)
<i>Xenopus</i>	xU1A	–ATTTG	(19)
Chicken	U1 2.5	–ATTTG	(10)
Chicken	U1 52a	–ATTTG	(10)
Rat	3-1A	–ATTTG	(44)
Rat	3-1B	–ATTTG	(44)
Mouse	U1B	–ATTTTTG	(32)

Sm antigen binding site, A(U)_nG, *n* = 3–6 (33).

similar sequence has been implicated as a protein binding site in human U1 (6, 23). Interestingly, the sequence A(U)_nG is also found at the immediate 3' terminus of the U1 coding sequence in *Xenopus* (8, 19), chickens (10), mice (32), and rats (44) (Table I). The conservation of this sequence through evolution gives further support to the hypothesis that the 3' tails of U1 precursors might have important functional roles.

The production of U1 transcripts from variant genes has several other important implications. Flanking sequences from human U1 true genes have been thought to be highly conserved. In fact, probes derived from the 5' flanking sequences of a prototype human U1 gene have been used to define the number and chromosomal location of active human U1 genes (3, 24–26, 38). If, as our data suggest, there are active genes with variant flanking sequences, it would be worthwhile to extend this analysis to include the flanking sequences from other active genes. Previous *in vitro* mutagenesis studies directed at characterizing the sequences responsible for the initiation and termination of U1 transcriptions have also been focused around the study of the flanking sequences from the prototype gene. Our data suggest that the conclusions drawn from these studies may not be applicable to all expressed U1 genes. Particularly interesting in this respect is the fact that at least one of the pseudogenes (pU1-6, reference 35), which has a 3' flanking sequence complementary to one of our oligonucleotide probes, has a true TATA box located at position –34 from the cap site. This raises the possibility that some functional U1 genes do contain the standard RNA polymerase II transcription signals.

As noted above, the expression of multiple U1 species seems to be a general phenomenon in vertebrates. Although no direct links have been established, it is tempting to speculate that this variability is significant in the regulation of tissue-specific or developmentally regulated splicing (7, 21, 27). There is little corresponding data concerning the possibility of heterogeneity in snRNP proteins, which have been characterized only as bands in one-dimensional acrylamide gels. In concert with our continuing studies of heterogeneity of snRNA expression, we have also begun to examine the structure and organization of genes for snRNP proteins (48). Ultimately we hope to combine the information obtained from these two lines of investigation to determine how U1 snRNP heterogeneity might relate to the function of this RNA-protein complex in splicing.

The authors wish to thank Dr. Michael Getz and Dr. Larry Pease for oligonucleotide synthesis. We also thank Dr. David Stanford for helpful discussions, Anne Rohleder for technical assistance, and Debbie Pearson for the preparation of this manuscript.

This work was supported by National Science Foundation grant DMB 8505716, and in part by the Mayo Foundation.

Received for publication 19 June 1986, and in revised form 13 October 1986.

References

1. Adams, S. P., K. S. Kavka, E. J. Wykes, S. B. Holder, and G. R. Galuppi. 1983. Hindered dialkylamino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. *J. Am. Chem. Soc.* 105:661-663.
2. Beaucage, S. L., and M. H. Caruthers. 1981. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* 22:1859-1862.
3. Bernstein, L. B., T. Manser, and A. M. Weiner. 1985. Human U1 small nuclear RNA genes: extensive conservation of flanking sequences suggests cycles of gene amplification and transposition. *Mol. Cell. Biol.* 5:2159-2171.
4. Black, D. L., B. Chabot, and J. A. Steitz. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell.* 42:737-750.
5. Brantlant, C., A. Krol, J. Ebel, E. Lazar, H. Gallinaro, M. Jacob, J. Sriwidada, and P. Jeanteur. 1980. Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res.* 8:4143-4154.
6. Brantlant, 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 (Eur. Mol. Biol. Organ.) J.* 10:1259-1265.
7. Breitbart, R. E., H. T. Nguyen, R. M. Medford, A. T. Destree, V. Mahdavi, and B. Nadal-Ginard. 1985. Intricate combinatorial patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. *Cell.* 41:67-82.
8. Ciliberto, G., R. Buckland, R. Cortese, and L. Philipson. 1985. Transcription signals in embryonic *Xenopus laevis* U1 RNA genes. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1537-1543.
9. Denison, R. A., and A. M. Weiner. 1982. Human U1 RNA pseudogenes may be generated by both DNA and RNA mediated mechanisms. *Mol. Cell. Biol.* 2:815-828.
10. Earley, J. M., K. A. Roebuck, and W. E. Stumph. 1984. Three linked chicken U1 RNA genes have limited flanking DNA sequence homologies that reveal potential regulatory signals. *Nucleic Acids Res.* 12:7411-7421.
11. Eliceiri, G., and M. Sayavedra. 1976. Small RNAs in the nucleus and cytoplasm of HeLa cells. *Biochem. Biophys. Res. Commun.* 72:507-512.
12. Forbes, D. J., M. W. Kirschner, D. Caput, J. E. Dahlberg, and E. Lund. 1984. Differential expression of multiple U1 small nuclear RNAs in oocytes and embryos of *Xenopus laevis*. *Cell.* 38:681-689.
13. Hernandez, N. 1985. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1827-1837.
14. Hinterberger, M., I. Pettersson, and J. A. Steitz. 1983. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. *J. Biol. Chem.* 258:2604-2613.
15. Kato, N., and F. Harada. 1985. New U1 RNA species found in friend SFV (spleen focus forming virus)-transformed mouse cells. *J. Biol. Chem.* 260:7775-7782.
16. Kinlaw, C. S., B. L. Roberson, and S. M. Berget. 1983. Fractionation and characterization of human small nuclear ribonucleoproteins containing U1 and U2 RNAs. *J. Biol. Chem.* 258:7181-7189.
17. Krainer, A. R., and T. Maniatis. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell.* 42:725-736.
18. Kramer, A., W. Keller, B. Appel, and R. Luhrmann. 1984. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell.* 38:299-307.
19. Krol, A., E. Lund, and J. E. Dahlberg. 1985. The two embryonic U1 RNA genes of *Xenopus laevis* have both common and gene-specific transcription signals. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1529-1535.
20. Kunkel, G. R., and T. Pederson. 1985. Transcription boundaries of U1 small nuclear RNA. *Mol. Cell. Biol.* 5:2332-2340.
21. Laski, F. A., D. C. Rio, and G. M. Rubin. 1986. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell.* 44:7-19.
22. Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz. 1980. Are snRNPs involved in splicing? *Nature (Lond.)* 283:220-224.
23. Liautard, J.-P., J. Sriwidada, C. Brunel, and P. Jeanteur. 1982. Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar domain of U1, U2, U4 and U5 small nuclear RNAs. *J. Mol. Biol.* 162:623-643.
24. Lindgren, V., L. B. Bernstein, A. M. Weiner, and U. Francke. 1985. Human U1 small nuclear RNA pseudogenes do not map to the site of the U1 genes in lp36 but are clustered in lq12-q22. *Mol. Cell. Biol.* 5:2171-2180.
25. Lund, E., and J. Dahlberg. 1984. True genes for human U1 small nuclear RNA. *J. Biol. Chem.* 259:2013-2021.
26. Lund, E., C. Bostock, M. Robertson, S. Christie, J. L. Mitchen, and J. E. Dahlberg. 1983. U1 small nuclear RNA genes are located on human chromosome I and are expressed in mouse-human hybrid cells. *Mol. Cell. Biol.* 3:2211-2220.
27. Lund, E., B. Kahan, and J. E. Dahlberg. 1985. Differential control of U1 small nuclear RNA expression during mouse development. *Science (Wash. DC)*. 229:1271-1274.
28. Madore, S., E. D. Wieben, and T. Pederson. 1984. Eukaryotic small ribonucleoproteins: anti-La human autoantibodies react with U1 RNA-protein complexes. *J. Biol. Chem.* 259:1929-1933.
29. Madore, S., E. D. Wieben, and T. Pederson. 1984. Intracellular site of U1 small nuclear RNA. *J. Cell. Biol.* 98:188-192.
30. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.
31. Manser, T., and R. F. Gesteland. 1982. Human U1 loci: genes for human U1 RNA have dramatically similar genomic environments. *Cell.* 29:257-264.
32. Marzluff, W. F., D. T. Brown, S. Lobo, and S.-S. Wang. 1983. Isolation and characterization of two linked mouse U1B small nuclear RNA genes. *Nucleic Acids Res.* 11:6255-6279.
33. Mattaj, J. W., and E. M. De Robertis. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell.* 40:111-118.
34. Melton, D. A., P. A. King, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035-7056.
35. Monstein, H. J., K. Hammarstrom, G. Westin, J. Zabielski, L. Philipson, and U. Pettersson. 1983. Loci for human U1 RNA. Structural and evolutionary implications. *J. Mol. Biol.* 167:245-257.
36. Mount, S. M., I. Pettersson, M. Hinterberger, A. Karmas, and J. A. Steitz. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell.* 33:509-518.
37. Myers, R. M., Z. Larin, and T. Maniatis. 1985. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science (Wash. DC)*. 230:1242-1246.
38. Naylor, S. L., B. U. Zabel, T. Manser, R. Gesteland, and A. Y. Sakaguchi. 1984. Localization of human U1 small nuclear RNA genes to band p36.3 of chromosome 1 by in situ hybridization. *Somatic Cell Mol. Genet.* 10:307-313.
39. Padgett, R. A., S. M. Mount, J. A. Steitz, and P. A. Sharp. 1983. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell.* 35:101-107.
40. Pederson, T. 1972. Chromatin structure and the cell cycle. *Proc. Natl. Acad. Sci. USA.* 69:2224-2228.
41. Rogers, J., and R. Wall. 1980. A mechanism for RNA splicing. *Proc. Natl. Acad. Sci. USA.* 77:1877-1879.
42. Roop, D. R., P. Kristo, W. E. Stumph, M. J. Tsai, and B. W. O'Malley. 1981. Structure and expression of a chicken gene coding for U1 RNA. *Cell.* 23:671-680.
43. Skuzeski, J., E. Lund, J. Murphy, T. Steinberg, R. Burgess, and J. Dahlberg. 1984. Synthesis of human U1 RNA. II. Identification of two regions of the promoter essential for transcription initiation of position +1. *J. Biol. Chem.* 259:9345-9352.
44. Watanabe-Nagasu, N., Y. Itoh, T. Tani, K. Okano, N. Koga, N. Okada, and Y. Ohshima. 1984. Structural analysis of gene loci for rat U1 small nuclear RNA. *Nucl. Acids Res.* 11:1791-1801.
45. Wieben, E. D., and T. Pederson. 1982. Small nuclear ribonucleoproteins of *Drosophila*: identification of U1 RNA-associated proteins and their behavior during heat shock. *Mol. Cell. Biol.* 2:914-920.
46. Wieben, E. D., S. Madore, and T. Pederson. 1983. U1 small nuclear ribonucleoprotein studied by in vitro assembly. *J. Cell Biol.* 96:1751-1755.
47. Wieben, E. D., S. Madore, and T. Pederson. 1983. Protein binding sites are conserved in small nuclear RNA U1 from insects and mammals. *Proc. Natl. Acad. Sci. USA.* 80:1217-1220.
48. Wieben, E. D., A. M. Rohleder, J. M. Nenninger, and T. Pederson. 1985. Evolutionary conservation of a human autoimmune nuclear antigen probed with a cDNA clone. *Proc. Natl. Acad. Sci. USA.* 82:7914-7918.
49. Winter, E., F. Yamamoto, C. Almoguera, and M. Perucho. 1985. A method to detect and characterize point mutations in transcribed genes: amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells. *Proc. Natl. Acad. Sci. USA.* 82:7575-7579.
50. Wood, W. I., J. Gitschier, L. A. Lasky, and R. M. Lawn. 1985. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA.* 82:1585-1588.
51. Yang, V. W., M. R. Lerner, J. A. Steitz, and S. J. Flint. 1981. A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proc. Natl. Acad. Sci. USA.* 78:1371-1375.
52. Zieve, G., and S. Penman. 1976. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell.* 8:19-31.