Volume 1 · Number 2 · May 1991
GENE EXPRESSION

Organization and transient expression of the gene for human U11 snRNA

Clemens Suter-Crazzolara and Walter Keller

Department of Cell Biology, Biocenter, University of Basel, Basel, Switzerland

The nucleotide sequence of U11 small nuclear RNA, a minor U RNA from HeLa cells, was determined. Computer analysis of the sequence (135 residues) predicts two strong hairpin loops which are separated by seventeen nucleotides containing an Sm binding site (AAUUUUUUGG). A synthetic gene was constructed in which the coding region of U11 RNA is under the control of a T7 promoter. This vector can be used to produce U11 RNA in vitro. Southern hybridization and PCR analysis of HeLa genomic DNA suggest that U11 RNA is encoded by a single copy gene, and that at least three genomic regions could be U11 RNA pseudogenes. A HeLa genomic copy of a U11 gene was isolated by inverted PCR. This gene contains the U11 RNA coding sequence and several sequence elements unique for the U RNA genes. These include a Distal Sequence Element (DSE, ATTTGCATA) present between positions −215 and −223 relative to the start of transcription; a Proximal Sequence Element (PSE, TTCACCTTTACCAAAAATG) located between positions −43 and −63; and a 3′box (GTTAGGCGAAATATTA) between positions +150 and +166. Transfection of HeLa cells with this gene revealed that it is functioning in vivo and can produce U11 RNA.

Small nuclear RNAs are metabolically stable RNAs present in all eukaryotic cells, which account for 1% of the total cellular RNA in higher eukaryotes (Weinberg and Penman, 1968). U1 to U6 RNAs are the most abundant U RNAs, with a concentration of 2×10^5–1×10^6 copies per cell, whereas minor U RNAs (U7 to U13) are present in concentrations of about 3×10^4 copies per cell. All U RNAs have several features in common (for review see Reddy and Busch, 1988): their 5′ terminus is capped with 2,2,7-trimethylguanosine (TMG; U6 has an γ-monomethyl phosphate-ester cap-structure; Singh and Reddy, 1989); they are located in the nucleus (U3, U8, and U13 are present in the nucleolus; Suh et al., 1986; Reddy et al., 1985; Tyc and Steitz, 1989); nucleotides are strongly modified in the 5′ half but not in the 3′ half of the U RNAs; and the U RNA genes are transcribed by RNA polymerase II (except for U6, which is synthesized by RNA polymerase III; Kunkel et al., 1986; Reddy et al., 1987).

The U RNAs are assembled into snRNP particles, which contain a group of seven common proteins and one or more type-specific proteins (for review see Lührmann, 1988). Autoantibodies from patients with systemic lupus erythematosus (SLE) were found to immunoprecipitate the snRNP particles (reviewed in Zieve and Sauterer, 1990). These antibodies are directed against the common snRNP proteins which bind the Sm-binding site with the consensus sequence $RA(U)_{4-6}GR$ (Jacob et al., 1984). Although U6 RNA lacks the Sm binding site, it is also im-

Received March 4, 1991; revision accepted April 3, 1991.

Correspondence: Walter Keller, Department of Cell Biology, Biocenter, University of Basel, CH-4056 Basel, Switzerland
Tel 41-61-267-2060 Fax 41-61-261-6760

munoprecipitated with the Sm antibody due to its association (by RNA–RNA base pairing) with U4 RNA in one particle (Bringmann et al., 1984; Hashimoto and Steitz, 1984). In addition, the anti-TMG antibody recognizes the TMG-cap structure (Lührmann et al., 1982).

U-snRNPs have been shown to be involved in the removal of introns from pre-mRNA (U1, U2, U4, U5, and U6; for review see Maniatis and Reed, 1987, and Steitz et al., 1988), in the processing of pre-rRNA (U3, U8, and U13; Kass et al., 1990) and in the 3′ processing of histone pre-mRNA (U7; Schaufele et al., 1986; Mowry and Steitz, 1987).

The genes for the major U RNAs are present at 5–40 copies per human genome (Lund and Dahlberg, 1984; Van Arsdell and Weiner; 1984, Bark et al., 1986), whereas U7, a minor RNA, is encoded by 5–10 genes in sea urchin (De Lorenzi et al., 1986) and at least one gene in the mouse (Gruber et al., 1991). Pseudogenes can be present at numbers 10 times higher (Manser and Gesteland, 1981) but have only been well characterized for some of the major U RNA genes (Denison and Weiner, 1982). The U RNA genes have several features in common: they are preceded by a distal sequence element (DSE) around position −220 and a proximal sequence element (PSE) around position −50. The latter is a TATA-box homologue unique for this class of genes (reviewed in Dahlberg and Lund, 1988). Termination of transcription occurs within a 3′ box (Hernandez and Weiner, 1986). The U6 gene, which is transcribed by RNA polymerase III, has a stretch of 5 T residues as terminator and an essential AT-rich element between its PSE and the start site of transcription (Lobo and Hernandez, 1989).

Here we report the elucidation of the sequence of U11, the RNA present in a minor snRNP (Krämer, 1987) that has been implicated in the 3′ processing of precursors to nuclear mRNAs (Christofori and Keller, 1988). A genomic copy of the U11 gene was isolated by the inverted polymerase chain reaction. Transfections of HeLa cells with this gene showed that it is efficiently expressed in vivo.

Materials and methods

Materials

Chemicals were obtained from Fluka (Switzerland), enzymes were from Biofinex (Switzerland), radiolabeled isotopes were purchased from Amersham, and TLC cellulose-plates from Merck. T1-, T2-, and P1-ribonucleases were obtained from Sigma.

Cloning and sequencing

All cloning procedures were carried out according to Sambrook et al., 1989. Constructs were sequenced on both strands with Sequenase (USB).

DNA and RNA sequence information was stored and analyzed with the help of the Genetics Computer Group programs (Devereux, 1990) and the EMBL database.

Sequence determination of U11 RNA. As a source for U11 RNA, a chromatographic fraction (Mono Q2, fraction 55, Christofori and Keller, 1988) was used. RNA from this fraction was purified by proteinase K treatment and phenol extraction (Humphrey et al., 1987) and 3′ end-labeled with [α-^{32}P]Cp and T4 RNA ligase (England and Uhlenbeck, 1978), followed by preparative gel electrophoresis (Frendewey and Keller, 1985). Chemical sequencing of RNA was carried out according to Conway and Wickens (1984). Primer extension sequencing was carried out essentially according to Solnick (1985) and Parker and Steitz (1987). After the reverse transcriptase reaction, all samples were subjected to RNAse A (0.5 μg/μl of reaction, 46°C for 10 minutes) and proteinase K treatment (Humphrey et al., 1987). This mixture was phenol- and chloroform-extracted, precipitated with ethanol, and separated by 10% denaturing PAGE. Chemical sequencing of full length cDNA was performed according to Maxam and Gilbert (1977).

The 3′ end of U11 RNA was determined by T1-ribonuclease digestion. Human U1, U2, U4, U5, U6, and U11 RNAs were 3′ end labeled with [α-^{32}P]Cp and T4 RNA ligase and gel-purified. Radiolabeled U RNAs (2000 cpm) were treated with 2 units T1-ribonuclease (Sigma; Filipowicz and Shatkin, 1983) in a final volume of 5 μl. The reaction mixture was incubated at 37°C for 45 minutes. Digestion products were analyzed by 20% denaturing PAGE.

The 3′ terminal nucleotide was determined by labeling of U11 RNA as described above and subsequent T2-ribonuclease treatment (Konarska et al., 1981). The reaction products were analyzed by one-dimensional TLC on a cellulose

plate in isobutyric acid/NH3/H_2O, pH = 4.3 (577:38:385).

To determine the 5′ terminal nucleotide, the method of Efstratiadis et al. (1977) was followed. Gel-purified U11 RNA was decapped with tobacco acid pyrophosphatase (BRL), treated with calf intestine phosphatase, and 5′ end labeled with [γ-^{32}P]ATP and polynucleotide kinase. The reaction product was recovered by 10% denaturing PAGE and subjected to P1-ribonuclease digestion (Konarska et al., 1981). Thin layer chromatography (Nishimura et al., 1967) of the reaction products was performed on cellulose plates in isobutyric acid/NH_3/H_2O, pH = 4.3 (577:38:385) in the first dimension and isopropanol/HCl/H_2O (70:15:15) in the second dimension.

Construction of Cs23 and Cs32. To construct a synthetic U11 RNA gene, ten DNA oligodeoxynucleotides were synthesized, encompassing both strands of the entire U11 coding sequence, the T7 promoter, and a DraI site. See Table 1.

The oligodeoxynucleotides were labeled with [γ-^{32}P]ATP and polynucleotide kinase, heated to 80°C for 5 minutes, annealed by slow-cooling to 30°C for 30 minutes, and ligated with T4 DNA ligase at room temperature for two hours. This cycle was repeated eight times. The reaction mixture was extracted with phenol and chloroform and precipitated with ethanol; the reaction product was purified by native 10% PAGE (Frendewey and Keller, 1985). This fragment was cloned with T4 DNA ligase into 100 ng HindII digested M13 mp18 DNA. The ligation mixture was used to transform JM103 cells, and phage DNA of several clones was sequenced. The Cs23 construct contains the in vitro assembled U11 gene, and T7 transcription of the DraI truncated construct gives rise to a sense U11 RNA preceded by 3 guanosines at its 5′ terminus.

Cs32 was prepared by PCR amplification (Innis and Gelfand, 1990) with 1 pg of Cs23 DNA as substrate. The primers (final concentration 2 μM; T7 promoter and DraI site underlined) were:

lu3 TTTAAAAGGGCTTCTGTCGTGAGT

lu4 CTTAAGCTTAATACGACTCACTATAGGG-AAAGGGCGCCGGGACCAACG.

Reaction products were isolated by 2.5% agarose gel electrophoresis and cloned into the HindII site of M13 mp18. The ligation mixture was used to transform JM103 cells. Recombinant phage DNA was sequenced, and one clone was named Cs32. T7 transcription of the DraI truncated Cs32 construct results in an antisense U11 RNA, preceded by 3 guanosines at the 5′ terminus.

In vitro synthesis of U11 RNA. Two μg of DraI truncated Cs23 or Cs32 construct were incubated in a total reaction volume of 20 μl, containing 0.01% Triton X-100, 40 mM Tris-HCl (pH = 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 10 mM DTT, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 50 μM UTP, 4 μl [α-^{32}P]UTP (20 mCi/ml), 1 μl RNA guard (35 units; Pharmacia) and 10 units T7- or T3-RNA polymerase (Hamm et al., 1987). This mixture was incubated at 37°C for 30 minutes, followed by phenol and chloroform extraction and ethanol precipitation. The RNA was purified by preparative 4% denaturing PAGE (Frendewey and Keller, 1985). One to two μg of RNA were synthesized, corresponding to 2–5 × 10^9 cpm.

Table 1. DNA oligodeoxynucleotides synthesized to construct a synthetic U11 RNA gene.

cs1	TAATACGACTCACTATAGGGAAAAAGGGCTTCTGT
cs2	GTCCCGGCGCCCTTTAAA
cs3	TTTAAAGGGCGCCGGGACCAACGATCACCAGC
cs4	TCCCTATAGTGAGTCGTATTA
cs5	CGTGAGTGGCACACGTAGGGCAACTCGATTGCTCT
cs6	GCGTGCGGAATCGACATCAAGAGATTTCGGAAGCA
cs7	TAATTTTTTGGTATTTGGGCAGCTGGTGATGGTTG
cs8	GCCCAAATACCAAAAAATTATGCTTCCGAAATCTC
cs9	TTGATGTCGATTCCGCACGCAGAGCAATCGAGTTG
cs10	CCCTACGTGTGCCACTCACGACAGAAGCCCTTTT

Underlining indicates T7 promoter and DraI site.

Southern- and Northern-hybridizations. For Southern blotting, 30 µg human placenta DNA was digested with several enzymes, phenol-extracted, ethanol-precipitated, separated by 1% agarose gel electrophoresis, and blotted to Zetaprobe membrane (Bio-Rad) according to the manufacturer's guidelines.

Northern blotting was carried out as follows: 10 µg RNA were separated by 10% denaturing PAGE and electroblotted to Genescreen membrane (DuPont) in 0.1 M Tris-HCl (pH = 7.8), 5 mM EDTA, and 50 mM sodium acetate at 150 mA for 10 hours. The filter was irradiated with UV light (260 nm) at 2000 µW/cm^2 for 75 seconds.

High stringency hybridization of Southern and Northern blots was carried out at 42°C for 16 hours in 5 × SSC, 5% SDS, 2.5% dextranesulphate, 1 × Denhardt solution (Sambrook et al., 1989), 7 µg/ml tRNA, 20 µg/ml poly A, 50 µg/ml heparin, 50% formamide, 50 mM $NaPO_4$, pH = 6.5. For low stringency hybridizations, formamide was omitted. Riboprobes had activities of at least 1 × 10^6 cpm/ml. The Southern filter was washed at room temperature in 1 × SSC, 0.2 µg/ml RNAse A for 2 hours and exposed for thirteen days. Northern filters were washed in 2 × SSC, 1% SDS for 1.5 hours and exposed for 16 hours.

Identification of U11 pseudogenes. Two oligodeoxynucleotides were used in a standard PCR amplification of genomic HeLa DNA (Innis and Gelfand, 1990):

ot1 AAGGGCTTCTGTCGTGA

ot2 GCGCCGGGACCAACGATCACCAGCTG.

Reaction products of 130 bp were purified by 2.5% agarose gel electrophoresis and subcloned in M13 mp18 (digested with HindII); the inserts of the resulting clones were sequenced.

Isolation of a genomic copy of the U11 RNA gene. Total genomic DNA (2 µg) was digested with several restriction enzymes HaeIII, DdeI, or HindII). Digested DNA was extracted with phenol and precipitated. Ligation was carried out in a volume of 400 µl at 14°C for 16 hours. The circularized DNA fragments were phenol-extracted and subsequently digested with either AflIII or PvuII. Both digests were mixed, phenol-extracted, and used for PCR amplification with the oligodeoxynucleotides:

ol2 CGGAATTCGTGCCACTCACGACAGAAGC

ol4 CCGGATCCGTGATCGTTGGTCCCGGCGC.

These oligodeoxynucleotides contain either an EcoRI or a BamHI site at the 5′ end.

The final concentration of each primer in the reaction was 2 µM, and the temperature cycles for PCR amplification (Innis and Gelfand, 1990) were as follows: 5 minutes at 94°C (prior to enzyme addition), 30 seconds at 94°C, 30 seconds at 42°C and 120 seconds at 72°C (15 cycles); 30 seconds at 94°C, 30 seconds at 60°C and 120 seconds at 72°C (20 cycles). The reaction products were digested with BamHI and EcoRI, purified by electrophoresis on a 1% agarose, and cloned into BamHI/EcoRI digested pUC18 vector. DNA of several clones was analyzed by sequence determination of both strands. Two of these constructs are named pU11.inv1 and pU11.inv2, originating from amplification of the initial HaeIII and DdeI digests respectively.

Based on the sequence information obtained from the pU11.inv1 construct, two oligodeoxynucleotides were synthesized:

om3 CGGATCCAAATTGTGTTGATAAAGATCTAC

om5 GGAAGCTTGGGCCCTTCCTGCTTTGTTGCTTAA.

These two oligodeoxynucleotides contain either a BamHI or a HindIII site at the 5′ end.

PCR amplification with these two primers (final concentration of 2 µM) was carried out in the presence of 1 µg untreated genomic HeLa DNA. The cycles were: 15 seconds at 95°C, 15 seconds at 60°C and 60 seconds at 72°C (15 times). The reaction product was digested with HindIII and BamHI prior to 1% agarose gel purification. The resulting fragment was cloned into the BamHI and HindIII sites of the Bluescript KS$^+$ vector, and both strands were sequenced with oligodeoxynucleotides spaced 200 bp along both strands of the gene. The final construct, which contains the U11 gene, is named pU11.

Construction of pU11a and pU11a.ori. A partial digest of pU11 was prepared with PvuII, which has three sites in this plasmid. One site is located in the coding sequence of U11 RNA (position 107, Fig. 1). The plasmid pU11 (1 µg) was incubated with 50 units PvuII at 37°C for 5, 10,

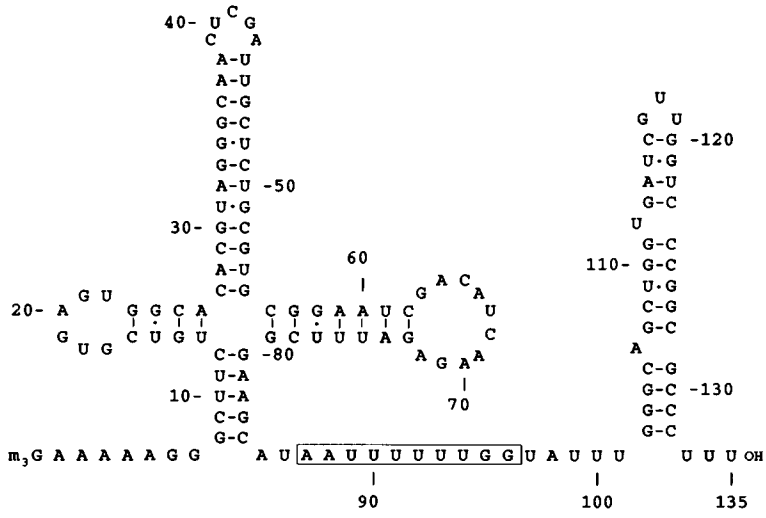

Figure 1. Sequence of U11 RNA. The secondary structure was predicted with the GCG "RNA Fold" program (Devereux, 1990). The Sm binding site (consensus $RAU_{[3-6]}GR$) is boxed, and the TMG cap is indicated as m_3G.

20, or 30 seconds; the reactions were stopped with phenol and combined. After phenol extraction and precipitation, linearized plasmid (digested only once by PvuII) was purified by 1% agarose gel electrophoresis.

Two complementary oligodeoxynucleotides:

alu-a CAGGTGATCCACCCACCTCGGCCTCC-CAAAGTG

alu-s CACTTTGGGAGGCCGAGGTGGGTGGA-TCACCTG

were mixed at equimolar concentrations (100 ng each), heated to 80°C, and cooled slowly to room temperature, resulting in a double-stranded fragment of 33 bp. This fragment was ligated into 100 ng PvuII linearized pU11. Clones containing the 33 bp fragment were isolated by colony hybridization with radiolabeled alu-a oligodeoxynucleotide. The sequence of positive clones was determined, and one plasmid which had the insert in the PvuII site of the U11 RNA coding region was named pU11a.

A SV40 origin of replication was inserted into the pU11a plasmid. The origin was obtained by PCR amplification of 1 pg pUC118/3 (Zenke et al., 1986; Hernandez and Lucito, 1988) with the M13 universal and reverse sequencing primers. Prior to 2.5% agarose gel purification, the 200 bp reaction product (containing the SV40 origin of replication and pUC118 polylinker sequences) was digested with HindIII and KpnI. The resulting fragment was cloned into the HindIII and KpnI sites of pU11a, resulting in the pU11a.ori plasmid.

Transfection of HeLa cells. Confluent HeLa cells were propagated in Dulbecco's Modified Eagle Medium with 5% calf serum (Gibco) and transfected by calcium-phosphate co-precipitation with the CellPhect Transfection Kit (Pharmacia). Twelve hours after transfection the medium was renewed, and 30 hours later total cellular RNA was isolated from 10^6 cells according to Summers (1970) and analyzed by Northern hybridization.

Results

The U11 RNA sequence

Several strategies were used to determine the nucleotide sequence of U11 RNA. First, U11 RNA was prepared from enriched chromatographic fractions of HeLa nuclear extracts (Christofori and Keller, 1988), labeled at the 3′ end with pCp, and partially chemically sequenced. Based on this information, an oligodeoxynucleotide complementary to nucleotides 101 to 128 of U11 RNA was synthesized. This oligodeoxynucleotide was used for primer extension sequencing and for the formation of a complete U11 cDNA, the latter being used as a substrate for chemical sequencing according to Maxam and Gilbert (1977). In order to obtain information on the 5′ terminal sequence of U11 RNA, primer extension sequencing was performed with an oligodeoxynucleotide complementary to nucleotides 24 to 66.

The 5′ and 3′ terminal nucleotides were determined as described under Materials and

Methods. The 5′ terminal nucleotide is a 2′-O-methyladenosine, whereas the 3′ terminal nucleotide is an unmodified uridine. The length of the 3′ terminal T1-digestion product was determined to be 6 nucleotides by comparison with T1-digestion products of several known U RNAs (data not shown). The complete sequence of U11 RNA is shown in Figure 1.

U11 RNA gene number

To estimate the number of genes that encode U11 RNA, a Southern blot of human placenta DNA digested with different restriction endonucleases was probed with radiolabeled synthetic U11 RNA (Fig. 2). The EcoRI, BamHI and NcoI digests each resulted in one strong signal (6500, 1600, and 3100 bp). For each digest, one band of weaker intensity (5-fold reduced) was also found (5300, 5000, and 2900 bp), as were very weak signals (at least 10-fold reduced; 2300 and 3000; 2300 and 8000; 2300 and 6000 bp). These weaker signals suggest that some of the fragments may not be completely homologous to the probe. These data indicate that the U11 gene is present as a single copy in the human genome.

Isolation of an U11 gene

An U11 gene was isolated by inverted PCR amplification (Triglia et al., 1988). HeLa DNA was digested with several restriction enzymes (HaeIII, DdeI, or HindII); the restriction fragments were cyclized and subsequently digested with PvuII and AflIII (both enzymes have sites within the U11 coding region). Theoretically, some resulting fragments contain parts of the U11 gene, with sequences of the U11 coding region at the 3′ and 5′ ends. These reaction products were subjected to PCR amplification with two oligodeoxynucleotides (ol2 and ol4) complementary to the 5′ and 3′ ends of the U11 coding region. These oligodeoxynucleotides point towards the center of the presumptive fragment. PCR products originating from the initial HaeIII, DdeI, and HindII digests (about 1000, 700, and 600 bp in length) were treated with BamHI and EcoRI and subcloned in pUC18. The sequence of the resulting constructs was then determined. Since the U11 gene has an internal EcoRI site, part of the upstream region is missing in these constructs. The pU11.inv1 and pU11.inv2 constructs originate from the 1000 and 700 bp fragments and contain over-

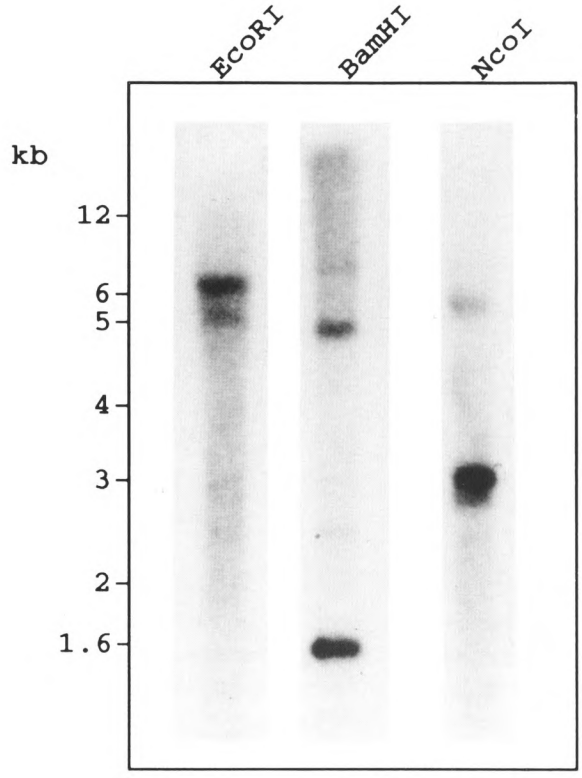

Figure 2. Southern analysis of human placenta DNA. Human placenta DNA was digested with EcoRI, BamHI, and NcoI; separated by 1% agarose gel electrophoresis; and transfected to Zetaprobe membrane. The filter was incubated with Cs32 riboprobe (containing U11 RNA preceded by 3 guanosine residues).

lapping, identical sequences. These include U11 specific sequences and sequences that are found in the upstream and downstream regions of other mammalian U RNA genes. Constructs that do not show these homologies were also identified: 30, 50, and 100% of the subclones resulting from the 1000, 700, and 600 bp fragments, respectively. Amplification of the initial DdeI digest also gave rise to fragments containing sequences with 70% homology to the U11 gene, as discussed below.

The construct pU11.inv1 contains part of the upstream and downstream regions of the presumptive U11 gene. Based on this information, two oligodeoxynucleotides (om3 and om5) were used to amplify the U11 gene by PCR from total genomic HeLa DNA. The resulting fragment (1130 bp) was subcloned into Bluescript (resulting in construct pU11) and its sequence was determined (Fig. 3).

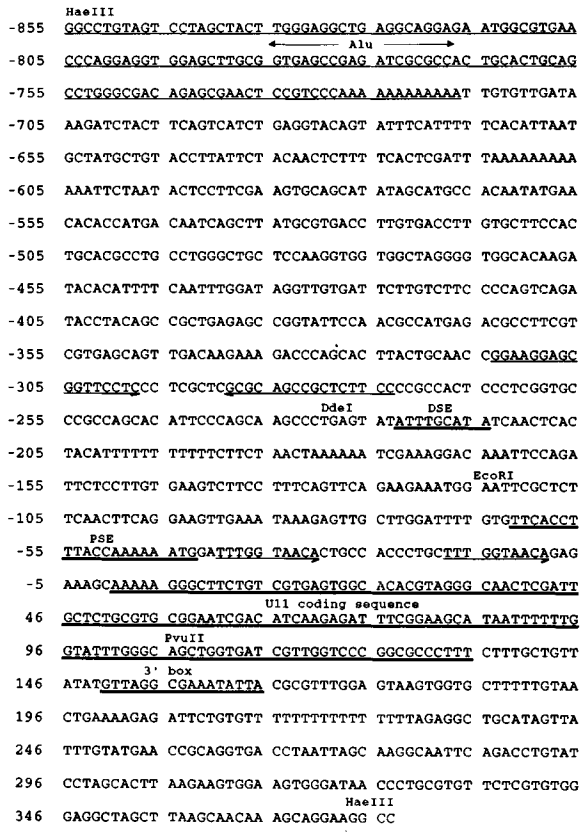

Figure 3. The human U11 gene. HeLa DNA digested with DdeI and HaeIII was used for cloning of the human U11 gene by inverted PCR. Shown here is the sequence of this locus between the two HaeIII sites. Thick underline: the U11 coding sequence (position +1 to +135), the distal sequence element (−223 to −215), the proximal sequence element (−62 to −41), and the 3′ box (+150 to +165). Further indicated are the Alu sequence (−855 to −717, thin underline); two indirect repeats and two direct repeats (arrows); and several restriction sites (HaeIII, GG′CC; DdeI, C′TGAG; EcoRI, G′AATTC; PvuII, CAG′CTG).

Putative U11 RNA pseudogenes

Three experiments indicate that U11 pseudogenes are present in the human genome. First, we carried out PCR amplification of the U11 coding region with two oligodeoxynucleotides located at the 3′ and 5′ ends of this region (ot1 and ot2). Reaction products of 130 bp were subcloned in M13 mp18. Six clones contained the true U11 coding sequence, whereas six others contained sequences which had only about 70% homology between the primers (sequence not shown).

A second indication for the presence of pseudogenes comes from analysis by Southern hybridization (Fig. 2). As described above, several signals with a low intensity were seen. These fall into two classes: one group of signals is 5-fold reduced in comparison to the strong signals (one band per digest); the other group contains signals which are at least 10-fold reduced (a minimum of two signals per lane).

Third, inverted PCR amplification of DdeI digested HeLa DNA resulted in the isolation of sequences that had homology with both U11 sequences and U RNA gene specific sequences (pU11.inv2). However, two of the six clones analyzed have only 68% sequence homology to pU11. inv1 (data not shown).

Characteristics of the U11 gene

Next to part of an Alu sequence (position −855 to −717, Fig. 3) is found a perfect octamer sequence (distal sequence element DSE, position −223 to −215), which functions as an enhancer in many genes (Mangin et al., 1986). In the HeLa U11 gene the orientation of this element is opposite to that in the human U1, U2, U3, and U4b genes, but the same as that in the human U6 and U4c genes (see legend of Figure 4 for references). A sequence motif between positions −62 and −43 shows large homology with the U RNA proximal sequence element (PSE); it fits the consensus sequence in 16 out of 20 positions (Fig. 4). The 3′ box consensus found downstream of most U RNA genes fits only partially to a region 3′ of the U11 gene, namely between positions 150 and 165. The sequence is homologous in 10 out of the 16 bp.

The A/T-rich motif found in the human U6 promoter is not present, nor is a stretch of five or more T residues directly at the end of the coding sequence of U6 (Lobo and Hernandez, 1989). This suggests that the U11 RNA gene is transcribed by RNA polymerase II, and not by RNA polymerase III.

Several continuous stretches of at least ten A or T residues are present in the genomic clone (positions 128 to 139, 242 to 253, 655 to 665, and 1069 to 1084). Also found are two inverted repeats, 17 bp in length with four mismatches (position 542 to 556, and position 569 to 582) and two direct repeats, 10 bp in length and completely homologous (position 816 to 825 and position 838 to 847). The latter repeats are located between the PSE and the start site of transcription.

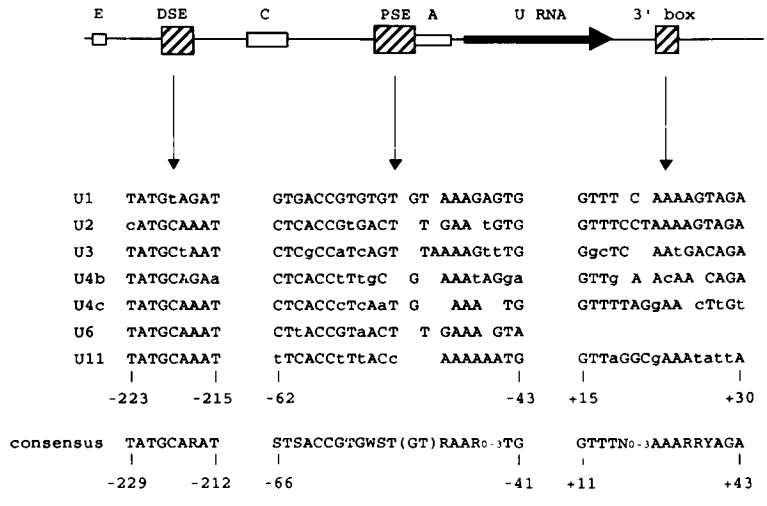

Figure 4. Comparison of the regulatory elements of human U RNA genes. The schematic (top) shows the general organization of human U RNA genes, including DSE, PSE, and 3′ box (Dahlberg and Lund, 1988). Below, these three transcription elements for each U snRNA gene are depicted in more detail. The consensus of each is shown at the bottom (R = A or G, Y = C or T, S = C or G and W = A or T). Residues in the transcription elements of the U snRNA genes that do not fit the consensus are shown in small print. The positions between which the consensus of each element is found are indicated by the numerals below (exception is the DSE of U4b, which is found around position +155). The positions of the three elements in the U11 gene are also indicated. In this figure, the position of the 3′ box is relative to the stop of transcription. Note: the U6 RNA gene is transcribed by RNA polymerase III and has a stretch of five T residues as terminator. The DSE elements of U4c, U6, and U11 are presented here in the opposite orientation of how they occur in their respective genes (see Murphy et al., 1982; Htun et al., 1985; Yuan et al., 1989; Bark et al., 1986; Kunkel et al., 1986).

Transient expression of the U11 gene

To determine whether the cloned U11 gene is expressed in vivo, the pU11 construct was first modified by insertion of a 33 bp fragment in the PvuII site present in the U11 coding region (construct pU11a). The tagged U11 RNA can easily be distinguished from the endogenous RNA by its difference in length. Furthermore, the SV40 origin of replication was inserted downstream of the U11 gene. This construct (pU11a.ori) was used to transfect HeLa cells. The plasmid pSVEori$^-$ (which encodes large T-antigen, Hernandez and Lucito, 1988) was always cotransfected to enable replication of constructs containing the SV40 origin.

Forty-two hours after transfection, total cellular RNA was isolated and analyzed by high stringency Northern blotting with Cs32 riboprobe. Such an experiment is shown in Figure 5A. Endogenous U11 RNA (135 nucleotides) is present in all transfected cells, whereas modified U11 RNA (168 nucleotides; U11a) is present only in cells transfected with pU11a.ori (lane 3 and 4), not in cells transfected with either pU11a (does not contain the SV40 origin of replication; lane 1 and 2) or pU2/−247 (contains a modified U2 gene, but no U11 related sequences; lane 5). This Northern filter was re-probed under low stringency conditions with a U2-specific probe. The Cs32 probe was not removed. Endogenous U2 RNA is present in all transfected cells, whereas the modified U2 RNA is found only in lane 5. The fact that the U2′ signal has a ten-fold lower intensity than the U11a signal is the result of the much higher stability of the U11a RNA (as observed in several experiments, data not shown). Due to the low stringency hybridization conditions, many RNAs crossreact with the probe, among which is probably 5S RNA. A U11 specific signal (marked "x" in Fig. 5A and B) could be a break-down product of both U11 and the modified U11a RNA, which is formed during RNA preparation.

Discussion

In this paper we report the complete nucleotide sequence of U11, the RNA moiety of a minor snRNP. This RNA was first identified by Krämer (1987). U11 RNA (Fig. 1) is 135 nucleotides long and shows some of the characteristics common to most U RNAs: it contains an Sm binding site (AAUUUUUUGG) which separates two putative hairpin-loop structures, and it is precipitable with anti-TMG antibodies, indicating that it carries a 2,2,7-trimethylguanosine cap structure. The sequence determined by us shows differences with the U11 RNA sequence previously reported by Montzka and Steitz (1988). The 5′ end of the sequence reported here is TMG-

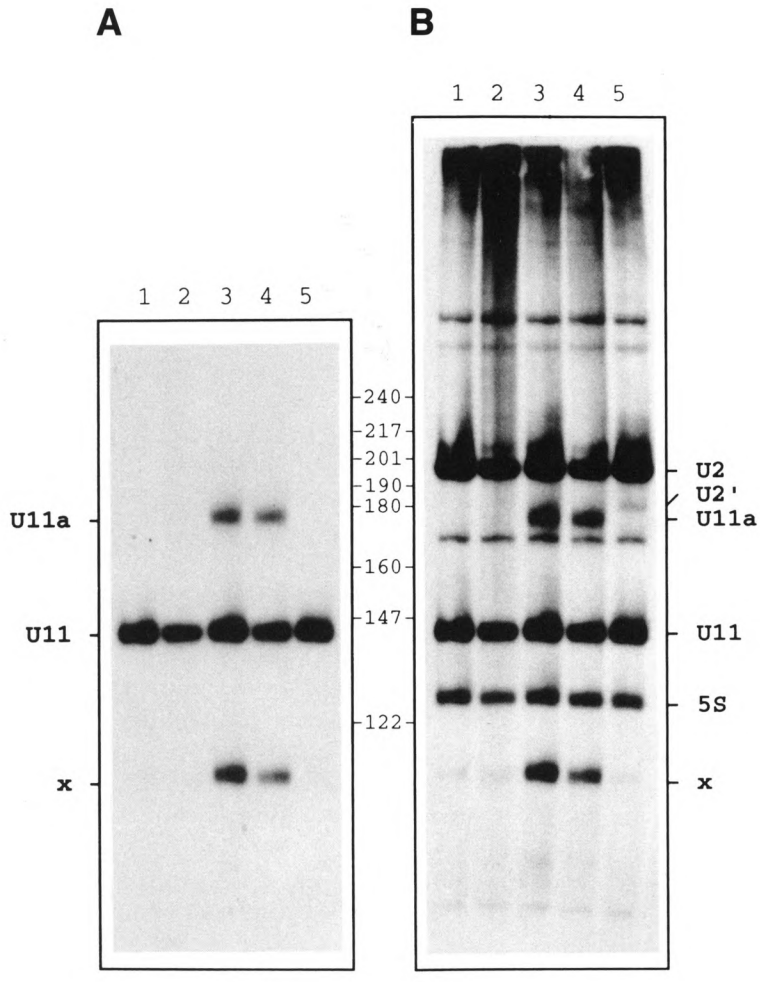

Figure 5. Transient expression of the U11 gene. HeLa cells were transfected by calcium phosphate co-precipitation. Twelve hours after transfection the medium was renewed, and total cellular RNA was isolated 30 hours later, separated by denaturing PAGE and analyzed by Northern hybridization. Cells were transfected with the following constructs: lane 1, 1.5 μg pU11a and 1.5 μg pSVEori⁻; lane 2, 3 μg pU11a and 3 μg pSVEori⁻; lane 3, 1.5 μg pU11a.ori and 1.5 μg pSVEori⁻; lane 4, 3 μg pU11a.ori and 3 μg pSVEori⁻; lane 5, 1.5 μg pU2/−247 and 1.5 μg pSVEori⁻. **A.** Northern filter probed with Cs32 riboprobe (high stringency conditions). **B.** the same filter probed under low stringency conditions with pU2/t3 (Cs32 riboprobe was not removed). Filters were exposed to film for 16 hours.

AAAAAGGGCU, whereas Montzka and Steitz report TMG-NAAGGCU. The 3′ end of the sequence shown here is CCCUUU-OH in contrast to CCCUU-OH. The fact that the sequence obtained from U11 RNA is identical to the U11 sequence in the genomic clone indicates that the sequence determination with U11 RNA is correct.

The only other minor U RNA of which the corresponding genes have so far been characterized is U7 RNA, which appears to be encoded in sea urchins by 5–10 genes, all present within a 9.3 kb fragment (De Lorenzi et al., 1986) and in mice by at least one gene (Gruber et al., 1991). Here we report on the genomic organization of the minor U11 RNA. From Southern hybridization (Fig. 2), we estimate that the U11 gene is present at 1 copy per haploid human genome. A more or less tight clustering of U RNA genes has been reported (reviewed by Dahlberg and Lund, 1988), which we cannot exclude for the U11 gene. However, the Southern blot shown in Figure 2 suggests that clustering of U11 genes is unlikely. This conclusion is based on the BamHI signal of 1600 bp, which should contain the 1232 bp fragment (shown in Fig. 3), leaving less than 368 bp for an additional U11 coding sequence. As is evident in Figure 2, at least three weaker bands are detected in all lanes of the Southern blot, indicating that other regions in the human genome have limited homology with the U11 RNA sequence. These regions could be pseudogenes. Indeed, both with PCR and inverted PCR, sequences were amplified independently, having about 70% homology to the U11 gene.

We have isolated a copy of a U11 gene from HeLa cells by inverted PCR amplification. This method was employed because conventional library screening did not result in the isolation

of a U11 gene. The disadvantage of this method is the rate of mis-incorporation by Taq polymerase. We tried to circumvent this problem by sequencing both the inserts of pU11.inv1 and pU11 completely. The first insert resulted from the inverted PCR amplification, the second from the PCR amplification. Furthermore, the upstream region of the U11 gene between the EcoRI site and the end of transcription was amplified by PCR separately and sequenced. Since no mismatches were found among pU11.inv1, pU11, and this additionally amplified product, we conclude that the sequence reported here (Fig. 3) is correct. The Alu sequence (position −855 to −717) is present only in pU11.inv1. Therefore, we cannot exclude that Taq polymerase-dependent misincorporation has taken place in this region.

Adjacent to the U11 coding sequence, this gene contains motifs that are typical for human U RNA genes (Fig. 4). Whereas the DSE sequence is completely homologous to the consensus sequence, the PSE and the 3′ box differ slightly from the consensus. The positions of these elements in the U11 gene are essentially identical to the positions in the other U RNA genes. Whether or not these sequence elements are required for transcription has not yet been determined.

The U11 gene is probably transcribed by RNA polymerase II. First, the TATA sequence found in the U6 gene between the PSE and the start of transcription is not present in the U11 gene. Second, the U11 coding sequence is not followed by a stretch of five or more T residues (a typical feature of RNA polymerase III-transcribed genes, as the U6 gene), but by a sequence that shows homologies to the 3′ box of U RNA genes. These data suggest that the human U11 gene described here belongs to the group of the RNA polymerase II-transcribed U RNA genes. Preliminary transfection experiments carried out in the presence of α-amanitin (which acts at low concentrations as an inhibitor of RNA polymerase II transcription; Dahlberg and Lund, 1988) support this hypothesis (data not shown).

The above implies that the U11 gene is a functioning gene. This was shown directly by transfection experiments of HeLa cells with the pU11a.ori construct (Fig. 5). A linker of 33 bp was inserted in the U11 coding sequence, increasing the length of U11 RNA to 168 n. Upon transfection of HeLa cells, this RNA (U11a) is formed in abundant amounts when the modified U11 gene is coupled to a SV40 origin of replication and large-T antigen is present (Fig. 5, lanes 3 and 4).

The U11 gene contains several features which are not present in other human U RNA genes (Fig. 3). First, 400 bp upstream of the DSE, an Alu sequence is found. We do not know whether this Alu element plays a role in transcription. However, pU11a.ori does not contain this element, and transcription in vivo is unaffected. We note that the human small RNA genes for RNAse P (Baer et al., 1990) and 7SK RNA (Murphy et al., 1986) have Alu elements within 200 bp downstream of their coding sequence.

Other novel features of the U11 gene are two indirect repeats and four stretches containing at least 11 A- or T- residues. One of the latter is located between the DSE and the PSE. Most striking however, are two identical direct repeats of 10 bp between the PSE and the start of transcription. We do not know what the possible functions of these sequence elements are. In this respect, it should be noted that U11 RNA is a minor RNA (Krämer, 1987), and thus present at concentrations a hundredfold lower than the major U RNAs. Whether this is the result of a low gene copy number, low transcriptional activity of the gene, or both is not known. However, from transfection experiments we have indications that U11 RNA is as stable as U1 RNA, suggesting that the lower concentration of U11 RNA is not regulated by rapid turn-over. The U11 gene does not contain additional enhancer elements. Putative SP1 sites, as found in the genes of U1, U2, and U3 (see legend of Fig. 4 for references) are absent from this gene.

The most important question remaining concerns the in vivo function of the U11 snRNP. Unfortunately, our results do not provide an answer. We have proposed previously that U11 snRNP may be involved in the 3′ processing of precursors to polyadenylated messenger RNAs. This assumption was based on the observation that U11 RNA was the only snRNA that cofractionated through several chromatographic steps, with the cleavage and polyadenylation factor CPF required for mRNA 3′ processing (Christofori and Keller, 1988). More recently it was reported that during the fractionation of CPF, most of the U11 RNA could be separated from CPF activity (Takagaki et al., 1989). We have confirmed these results, and we have es-

tablished a purification procedure which essentially removes all U11 snRNP from the cleavage and polyadenylation factor (S. Bienroth, C. Suter-Crazzolara, E. Wahle, and W. Keller, in preparation). Thus it appears that U11 snRNP is not essential for in vitro 3′ processing and could represent a contaminant in partially purified CPF preparations.

Attempts to isolate genes coding for an U11 RNA homologue in yeast have so far not been successful. This is unfortunate, since disruption of the gene in yeast and reconstitution with plasmids carrying the U11 gene under the control of an inducible promoter would probably allow testing for the physiological role of U11 snRNP.

Acknowledgments

We are grateful to Laura Conway for part of the chemical sequence analysis of U11 RNA. We thank Witek Filipowicz and Hans Gross for their advice and markers for the TLC analysis. We thank Nouria Hernandez for pUC118/3, pU2/−247, pSVEori⁻, and pU2/t3; Graeme Bilbe for the Southern filter used for Fig. 3; Reinhard Lührmann for the anti-TMG antibody; and Iain Mattaj for his help with the synthetic U11 construct.

This work was supported by grants from the Schweizerischer Nationalfonds and the Kantons of Basel.

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession number X58716.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

Clemens Suter-Crazzolara is currently at the MPI für Züchtungsforschung, Carl von Linné Weg 10, 5000 Köln 30, Germany.

References

M. Baer, T. W. Nilsen, C. Costigan, and S. Altman (1990), Nucl Acids Res 18, 97–103.

C. Bark, P. Weller, J. Zabielski, and U. Pettersson (1986), Gene 50, 333–344.

P. Bringmann, B. Appel, J. Rinke, R. Reuter, H. Theissen, and R. Lührmann (1984), EMBO J 3, 1357–1363.

G. Christofori and W. Keller (1988), Cell 54, 875–889.

L. Conway and M. Wickens (1984), EMBO J 6, 4177–4184.

J. E. Dahlberg and E. Lund (1988), in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles (M. L. Birnstiel, ed.), Springer-Verlag, Berlin and New York, pp. 38–70.

M. De Lorenzi, U. Rohrer, and M. L. Birnstiel (1986), Proc Natl Acad Sci USA 83, 3243–3247.

R. A. Denison and A. M. Weiner (1982), Mol Cell Biol 2, 815–828.

J. Devereux (1990), Nucl Acids Res 12, 387–395.

A. Efstratiadis, J. N. Vournakis, H. Donis-Keller, G. Chaconas, D. K. Dougall, and F. C. Kafatos (1977), Nucl Acids Res 4, 4165–4174.

T. E. England and O. C. Uhlenbeck (1978), Nature 275, 560–561.

W. Filipowicz and F. Shatkin (1983), Cell 32, 547–557.

D. Frendewey and W. Keller (1985), Cell 42, 355–367.

A. Gruber, D. Soldati, M. Burri, and D. Schümperli (1991), Biochim Biophys Acta 1088, 151–154.

J. Hamm, M. Kazmaier, and I. W. Mattaj (1987), EMBO J 6, 3479–3485.

C. Hashimoto and J. A. Steitz (1984), Nucl Acids Res 12, 3283–3293.

N. Hernandez and R. Lucito (1988), EMBO J 7, 3125–3134.

N. Hernandez and A. M. Weiner (1986), Cell 47, 249–258.

H. Htun, E. Lund, G. Westin, U. Petterson, and J. E. Dahlberg (1985), EMBO J 4, 1839–1845.

T. G. Humphrey, G. Christofori, V. Lucijanic, and W. Keller (1987), EMBO J 6, 4159–4168.

M. A. Innis and D. H. Gelfand (1990), in PCR Protocols (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds.), Academic Press, New York, pp. 3–13.

M. Jacob, E. Lazar, B. Haendler, H. Gallinaro, A. Krol, and C. Branlant (1984), Biol Cell 51, 1–10.

S. Kass, K. Tyc, J. A. Steitz, and B. Sollner-Webb (1990), Cell 60, 897–908.

M. Konarska, W. Filipowicz, H. Domdey, and H. J. Gross (1981), Nature 293, 112–116.

A. Krämer (1987), Proc Natl Acad Sci USA 84, 8408–8412.

G. R. Kunkel, R. L. Maser, J. P. Calvet, and T. Pederson (1986), Proc Natl Acad Sci USA 83, 8575–8579.

S. M. Lobo and N. Hernandez (1989), Cell 58, 55–67.

E. Lund and J. E. Dahlberg (1984), J Biol Chem 259, 2013–2021.

R. Lührmann (1988), in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles (M. L. Birnstiel, ed.), Springer-Verlag, Berlin and New York, pp. 71–99.

R. Lührmann, B. Appel, P. Bringmann, J. Rinke, R. Reuter, S. Rothe, and R. Bald (1982), Nucl Acids Res 10, 7103–7113.

M. Mangin, M. Ares, Jr., and A. Weiner (1986), EMBO J 5, 987–995.

T. Maniatis and R. Reed (1987), Nature 325, 673–678.

T. Manser and R. F. Gesteland (1981), J Mol Appl Genet 1, 117–125.

A. Maxam and W. Gilbert (1977), Proc Natl Acad Sci USA 74, 560–564.

K. A. Montzka and J. A. Steitz (1988), Proc Natl Acad Sci USA 85, 8885–8889.
K. L. Mowry and J. A. Steitz (1987), Science 238, 1682–1687.
J. T. Murphy, R. R. Burgess, J. E. Dahlberg, and E. Lund (1982), Cell 29, 265–274.
S. M. Murphy, A. Tripodi, and M. Melli (1986), Nucl Acids Res 14, 9243–9260.
S. Nishimura, F. Harada, U. Narushima, and T. Seno (1967), Biochim Biophys Acta 142, 133–148.
K. A. Parker and J. A. Steitz (1987), Mol Cell Biol 7, 2899–2913.
R. Reddy and H. Busch (1988), in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles (M. L. Birnstiel, ed.), Springer-Verlag, Berlin and New York, pp. 1–35.
R. Reddy, D. Henning, and H. Busch (1985), J Biol Chem 260, 10930–10935.
R. Reddy, D. Henning, G. Das, M. Harless, and D. Wright (1987), J Mol Biol 262, 75–81.
J. Sambrook, E. F. Fritsch, and T. Maniatis (1989), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
F. Schaufele, G. M. Gilmartin, W. Bannwarth, and M. L. Birnstiel (1986), Nature 323, 777–780.
R. Singh and R. Reddy (1989), Proc Natl Acad Sci USA 86, 8280–8283.
D. Solnick (1985), Cell 43, 667–676.
J. A. Steitz, D. Black, V. Gerke, K. A. Parker, A. Krämer, D. Frendewey, and W. Keller (1988), in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles (M. L. Birnstiel, ed.), Springer-Verlag, Berlin and New York, pp. 115–154.
D. Suh, H. Busch, and R. Reddy (1986), Biochim Biophys Res Commun 137, 1133–1140.
W. C. Summers (1970), Anal Biochem 33, 459–463.
Y. Takagaki, L. Ryner, and J. L. Manley (1989), Genes Dev 3, 1711–1724.
T. Triglia, M. G. Peterson, and D. J. Kemp (1988), Nucl Acids Res 16, 8136.
K. Tyc and J. A. Steitz (1989), EMBO J 8, 3113–3119.
S. W. Van Arsdell and A. M. Weiner (1984), Mol Cell Biol 4, 492–499.
R. Weinberg and S. Penman (1968), J Mol Biol 38, 289–304.
Y. Yuan and R. Reddy (1989), Biochim Biophys Acta 1008, 14–22.
M. Zenke, T. Grundström, H. Matthes, M. Wintzerith, C. Schatz, A. Wildemann, and P. Chambon (1986), EMBO J 5, 387–393.
W. G. Zieve and R. A. Sauterer (1990), Crit Rev Biochem Mol Biol 25, 1–46.

Note added in proof

A YGTGTTYY motif, located downstream of the poly(A) addition site, is required for the efficient formation of mRNA 3′ termini (J. McLauchlan, D. Gaffney, J. L. Whitton, and J. B. Clements [1985] Nucl Acids Res 13, 1347–1368). By sequence comparison of 47 precursor mRNAs, we conclude that the consensus of this element in humans is TGTGYYYY. We observe that nucleotides 21 to 28 of U11 RNA have complementarity to this sequence element (allowing for G·U base pairing). Thus, it is possible that U11 snRNP may be involved in the cleavage reaction preceding polyadenylation by binding to this element. Complete purification of all the cleavage factors that participate in 3′ processing will be necessary to settle this issue.