The human U1 snRNA promoter correctly initiates transcription in vitro and is activated by PSE1

Samuel I. Gunderson, Mark W. Knuth, and Richard R. Burgess
McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706 USA

A DNA-dependent in vitro transcription system for the human U1 small nuclear RNA (snRNA) promoter has been developed. This in vitro transcription system uses extracts of tissue culture cells to drive transcription of an RNA polymerase II-transcribed snRNA gene. A U1 promoter (−393 to +192) template was constructed in which the sequences from +10 to +171 were replaced with a 179-bp sequence from a G-less cassette. This DNA template thus retained all of the known U1 promoter elements, including the U1 3'-end box (positions +175 to +191), which is responsible for snRNA 3'-end formation. HeLa cell nuclear extracts were shown to drive specific transcription of this promoter by RNA polymerase II. This transcription system has many of the properties observed for wild-type snRNA promoters in vivo. Transcription was shown to initiate at +1 (and −2) relative to the U1 promoter and to efficiently (＞90%) form a 3’ end corresponding to the 3’ end found in the primary transcript of U1 in vivo. The transcription signal is responsive to either deletion or replacement of the U1 distal sequence (enhancer-like) and proximal sequence (TATA-like) elements, as well as the 3'-end box. Additionally, the signal was shown by depletion/repletion experiments to be responsive to a protein called PSE1 (related to Ku), which has recently been shown to specifically bind sequences in the U1 promoter. This in vitro snRNA transcription system should facilitate the biochemical analysis of the human U1 snRNA promoter and lead to a better understanding of the differences between snRNA and mRNA promoters.

[Key Words: Small nuclear RNA; U1; in vitro transcription; 3'-end formation; G-less cassette]

Received June 20, 1990; revised version accepted October 2, 1990.

Eukaryotic small nuclear RNAs (snRNAs) are synthesized by RNA polymerase II [Frederickson et al. 1978], with the exception of U6, which is transcribed by RNA polymerase III [Kunkel et al. 1986]. In spite of sharing a common RNA polymerase, RNA polymerase II-transcribed snRNA promoters are distinct from mRNA promoters [for review, see Dahlberg and Lund 1988] in the following characteristics: (1) The position of the start site is determined by a snRNA-specific proximal sequence element [PSE] located 40–80 bp upstream of the initiation site; there is no apparent TATA box at −30 (Skuzeski et al. 1984; Mattaj et al. 1985). [2] 3'-End formation is directed by a snRNA-specific element called the 3'-end box, which is found 10–30 bp downstream of the coding region [Hernandez 1985; Yuo et al. 1985]. [3] 3'-End formation is tightly coupled to initiation directed by snRNA promoters. That is, transcripts that initiate distantly from +1 do not terminate under the direction of the 3'-end box, but instead utilize downstream mRNA 3'-end signals [the AAUAAA sequence] [Hernandez and Weiner 1986; Neuman de Veygar et al. 1986]. [4] snRNA promoter elements are not functionally interchangeable with mRNA promoter elements [Hernandez and Weiner 1986; Neuman de Veygar et al. 1986; Dahlberg and Schenborn 1988]. [5] Transcription initiation of snRNAs occurs at a single site [unless heterologous systems are used], whereas heterogenous starts are commonly observed for mRNA promoters [Dahlberg and Lund 1988]. Three major functional elements are thought to be common to all vertebrate snRNA genes. The distal sequence element [DSE], located ~200–250 bp upstream of the initiation site (+1), modulates transcription activity in vivo by one to two orders of magnitude [Skuzeski et al. 1984; Westin et al. 1984; Mattaj et al. 1985; Murphy et al. 1987]. The DSE is often called the snRNA enhancer because it functions independently of orientation, however, unlike classical enhancer elements, its activity exhibits distance dependence, and it is inactive when placed downstream of the coding region [Mattaj et al. 1985; Murphy et al. 1987]. The PSE, located 40–80 bp upstream from the initiation site, is absolutely required for transcription activity in vivo. Because the PSE positions the location of the initiation site but does not contain a TATA-like sequence, it is sometimes referred to as the snRNA TATA equivalent [Skuzeski et al. 1984; Mattaj et al. 1985; Murphy et al. 1987]. Vertebrate PSES contain a conserved core sequence that has been shown to be critical for promoter activity [Hernandez and Lucito 1988; Parry et al. 1989]. The third element, the 3’-
end box, located 10–30 bp downstream of the coding region, directs 3'-end formation. When the 3'-end box is deleted, transcription complexes initiating at +1 will utilize cryptic 3'-end boxes, even when mRNA termination signals are present (Hernandez 1985; Yuo et al. 1985; Hernandez and Weiner 1986; Neuman de Vegvar et al. 1986).

In the cell, snRNA genes are extremely active promoters. For each of the 40–60 human U1 genes, a transcript is initiated every 2–4 sec. Paradoxically, whole-cell or nuclear extracts competent for transcription of mRNA genes have not been able to synthesize correctly initiated snRNA products, even when very sensitive detection methods are employed (for review, see Dahlberg and Lund 1988). Two in vitro transcription systems for snRNA promoters have recently been developed. In one system, extracts from sea urchin embryos were used to direct synthesis of sea urchin snRNAs (Morris et al. 1986). However, sea urchin snRNAs (and other invertebrate snRNAs) use different transcription signals than vertebrate snRNAs (Saba et al. 1986; Strub and Birnstiel 1986). Thus, the sea urchin system has not led to success for vertebrate snRNAs. More recently, an in vitro transcription system using *Xenopus laevis* oocyte germinal vesicles (GVs) was reported (Lund and Dahlberg 1989), which faithfully directs +1 initiation and 3'-end formation of vertebrate snRNAs (*Xenopus* and human). In this system, the GVs were isolated individually under mineral oil, transferred to an aqueous buffer, and homogenized.

Although the human U1 gene is transcriptionally active in the GV system, we wanted to study the human U1 gene in a completely homologous system using extracts of human tissue culture cells. In this paper we report the development of a DNA-dependent in vitro transcription system using HeLa cell nuclear extracts for the human U1 promoter fused to the G-less (G~) cassette (Sawadogo and Roeder 1985). RNA polymerase II-directed transcription was shown to initiate at +1 and to form a 3' end at a position corresponding to that found when the U1 gene was injected into *X. laevis* oocytes (Neuman de Vegvar et al. 1986). In vitro transcription activity was dependent on promoter elements previously shown to be important to the synthesis of the human U1 gene in vivo.

Finally, we demonstrate by immunodepletion/repletion experiments that PSE1, a transcriptional activating protein that specifically binds sequences in the PSE and DSE of the U1 promoter (Gunderson et al. 1988; Knuth et al. 1990), is involved in the transcription of the U1G~ system. This underscores the potential utility of the in vitro transcription system for the identification and purification of other U1 transcription factors.

**Results**

**Preliminary attempts**

In attempting to develop an in vitro transcription system, a variety of extracts were prepared and tested on the wild-type U1 gene using sensitive transcription assays, including nuclease protection assays (using uniformly labeled probes) and hybrid selection [to pull out of solution transcripts containing U1-coding sequences]. All of these approaches failed to detect transcripts initiating at +1. In a different approach, a standard 25-μl transcription reaction was scaled up 30-fold, and the RNA was extracted and size-fractionated by preparative denaturing gel electrophoresis. We were able to demonstrate by primer extension that one of the gel-purified RNA fractions had a 5' end corresponding to +1 initiation (S. Gunderson, unpubl.). However, this potential +1 signal [even at 30-fold scale-up] was extremely faint. We abandoned this approach because the assay was time-consuming, hard to reproduce, and consumed large quantities of extract and plasmid DNA. Because we did detect a +1 start signal, we searched for alternative approaches.

**The G~ cassette assay**

The G~ cassette assay was developed by Sawadogo and Roeder (1985) as a rapid and facile assay to analyze RNA products from in vitro transcription reactions. The major advantage of this assay is that it suppresses transcriptional activity from cryptic mRNA-like promoters found within the U1 promoter. The RNA chain-terminating nucleotide, 3'-O-methyl-GTP, terminates RNA polymerase elongation complexes prematurely. Because nuclear extracts have a low level of endogenous nucleotides [10–20 μM], some transcription complexes will read through the guanosines, giving rise to readthrough transcripts (for detailed discussion, see Fig. 2A). RNase T1 is therefore included in the transcription reaction to trim these readthrough transcripts to a defined length.

**Construction and description of the U1G~ and HTFRG~ promoters**

Figure 1 summarizes the structure of the promoter constructs used in this paper. All of the mutated U1G~ constructs are versions of pSG611, which represents the parental U1G~ promoter. For the remainder of this paper, the trivial names of the promoter templates will be used instead of the formal construct names. The U1G~ promoter has several features that are crucial to understanding the results presented in this paper. First, U1G~ retains all of the known U1 promoter elements, which include the DSE [or snRNA enhancer], the PSE, the sequences around the initiation site, and the 3'-end box [the element for snRNA 3'-end formation]. Second, although U1G~ lacks the U1 initiation site sequence from +9 to +164, it has been shown that snRNA-coding regions are dispensable for snRNA-type transcription (Dahlberg and Lund 1988; Dahlberg and Schenborn 1988). Third, the region lacking Gs [the G~ region] extends from −5 to +188 and incorporates the U1 initiation site sequences from −5 to +9. Fourth, the U1 3'-end box is placed immediately adjacent to the G~ region so that the 3' ends of transcripts formed by this element are distinct from the 3' ends of mRNA-like transcripts. Fifth, some transcripts that initiate upstream of −6 are expected to read
Figure 1. Schematic of promoter constructs. The structures of the UIG- and HTFRG- promoter constructs used in this paper are diagramed. [Top] Arrows denote both the boundaries and the origins of the sequences contained in the constructs. [Left] Construct names and trivial names used throughout this paper. All of the constructs share a common core sequence, the initiation/transcribed region, extending from -7 to +191. The sequences from -7 to +9 are from the Ul gene (Ul sequences -7 to +9); the sequences from +10 to +188 are from the upstream portion of the G- cassette (G- sequences from +10 to +188; see Sawadogo and Roeder 1985). The GAG sequence at +189 to +191 produces an Aval site and came from subcloning. As indicated by the arrows, the parent construct pSG611 places the Ul upstream promoter (-393 to -8) in front of the -7 to +191 region and places a Ul 3'-end box (Ul sequences from +172 to +191) immediately downstream of the same -7 to +191 region at the Aval site. Construct pSG628 places a HTFR promoter (from -100 to -8) upstream of the -7 to +191 region. The predicted 3' end of HTFRG- synthesized RNA (pSG628) is also indicated. The sequences of the oligonucleotides used to make mutated promoters pSG638 and pSG636 are given in Materials and methods; and in both promoters, the oligonucleotide maintained the same spacing between the promoter elements as is found in pSG611.

through the upstream guanosines and traverse the G- cassette. RNase T1 is expected to trim these transcripts to a length of 194 nucleotides corresponding to -5 to +189. These transcripts are designated readthrough transcripts [discussed in Fig. 2A, below] although they would be the same length as transcripts that initiate at -5 and terminate at +189.

The HTFRG- template, pSG628, replaces U1 sequences -393 to -8 in the UIG- template with a 92-bp sequence from the human transferrin receptor (HTFR) promoter corresponding to the minimal HTFR promoter from -100 to -8 (Casey et al. 1988). The HTFRG- template contains a TATA box at -30 and initiates transcription at the U1 +1 nucleotide [see Fig. 4]. HTFRG- transcribed RNA serves as a useful size marker in our denaturing gels because it should migrate as a +1 initiated transcript. We did not expect HTFRG- derived transcription complexes to use the U1 3'-end box, and deletion of the 3'-end box of HTFRG- does not affect transcriptional activity (data not shown). Thus, HTFRG- represents mRNA-like transcription that should initiate at +1 and terminate at position +189 due to incorporation of 3'-O-methyl-GTP.

Initial detection of and improvement in strength of the UIG- transcription signal

The UIG- transcription signal initially obtained was extremely weak, only being visible on a 10- to 12-day autoradiographic exposure with an intensifying screen (data not shown). Prior to examining the properties of this signal, transcription reaction parameters were therefore optimized to maximize signal strength. Details of the optimization are found in Materials and methods and in Figure 8. Although other parameters had significant effects, the UIG- +1 transcription activity was most sensitive to relatively small changes in MgCl2, with an optimal value of 3.5 mM. The optimized conditions resulted in a 10-fold increase in signal strength, making characterization of its properties much easier. The HTFRG- promoter transcription conditions were also optimized (data not shown).

Effects of 3'-O-methyl-GTP and RNase T1 on UIG- and HTFRG- transcription

To assess the degree to which 3'-O-methyl-GTP sup-
presses spurious transcription activity and RNase T1 degrades guanosine-containing transcripts, these two components were varied across a wide range of concentrations. Figure 2A illustrates which transcripts are predicted to form under low or no 3'-O-methyl-GTP conditions, both in the presence and absence of RNase T1. Figure 2B shows what type of transcripts arise when these two components were varied for the U1G- (lanes 1–6) and HTFRG- (lanes 7–11) promoters. When both components were omitted (lanes 1 and 7), transcription complexes elongated around the plasmid until they reached ill-defined pause sites or termination signals. Thus, this produced transcripts of all sizes, visible as a smear on the autoradiogram. When 0.1 mM 3'-O-methyl-GTP alone was included (lanes 2 and 8), the background transcription signal was suppressed somewhat and the HTFRG- +1 [-1 to +189] was visible. For U1G- the level of background from readthrough transcription was still too high to see the U1-specific starts. When RNase T1 alone was included (data not shown), all transcripts spanning the G- region were trimmed to the length of the G- region corresponding to position -5 to +189 (see Fig. 1). The use of RNase T1 alone therefore resulted in a massive readthrough transcript signal being observed (data not shown).

When 0.1 mM 3'-O-methyl-GTP was used and RNase T1 was added during the last 20 min of transcription (Fig. 2B, lanes 3 and 9), the background level of transcripts was lowered such that U1-specific bands could be seen; these are labeled as -2 and +1 starts in Figure 2B. Under these conditions, the HTFR +1 signal increased ~10-fold (cf. lanes 8 and 9). This increase implies two things: [1] The HTFR transcription complex was reading through the U1 3'-end box, and [2] 0.1 mM 3'-O-methyl-GTP was not high enough to terminate transcription efficiently at every guanosine (~10% efficiency). The appearance of the readthrough transcript in lane 3 further supports the fact that 0.1 mM 3'-O-methyl-GTP was not completely suppressing upstream transcription. Band patterns identical to lanes 3 and 9 were obtained if α-amanitin was added along with RNase T1 at 70 min, demonstrating that RNase T1 was not inducing the production of the U1- or HTFR-specific +1 transcripts (data not shown).

When 0.8 mM 3'-O-methyl-GTP was used without RNase T1 (Fig. 2B, lanes 4 and 10), transcription was terminated efficiently at every guanosine (>90%), and background from readthrough transcription was greatly suppressed. This eightfold increase in 3'-O-methyl-GTP did not appreciably affect the accumulation of the +1

---

**Figure 2.** Effects of 3'-O-methyl-GTP and RNase T1 on the U1G- and HTFRG- transcription signals. **A** The schematic illustrates what kind of HTFRG- and U1G- transcripts are synthesized under low or no 3'-O-methyl-GTP conditions. Indicated at the top are the U1G- and HTFRG- promoters, both of which are predicted to act as templates for synthesis of three types of transcripts: [1] Readthrough transcripts (both HTFRG- and U1G-), [2] +1-initiated transcripts that are independent of the 3'-end box (HTFRG- and perhaps U1G-), and [3] +1-initiated transcripts that are dependent on the 3'-end box ([U1G- only]). The initial transcript is indicated by a heavy solid line with an arrowhead. This transcript is then trimmed by RNase T1 to form the final transcript, which is indicated by a hatched box. The question mark (!) serves to indicate that the location of the 3'-end cannot be predicted precisely when 3'-end formation is under the direction of the 3'-end box. **B** Standard transcription reaction conditions were used except that 3'-O-methyl-GTP and RNase T1 were varied for the entire reaction and was at the concentration indicated above each lane. RNase T1 was present only in lanes 3, 5, 9, and 11, where it was added only after 70 min had elapsed in the 90-min transcription reaction. The position of endogenous U1 is indicated and arose by 3'-end addition of [32P]UTP due to the activity of endogenous enzymes in the extract. The endogenous U1 band is more visible in lanes 7, 8, and 10 after a 48-hr exposure. Numbers at right indicate the position and lengths of DNA markers described in Fig. 3. The control RNA (also described in Fig. 3) is not shown but varied less than twofold among lanes. Autoradiographic exposure times with an intensifying screen are indicated.
start signals for either UIG- or HTFRG- (cf. lanes 3 with 4 and 9 with 10). When 0.8 mM 3'-O-methyl-GTP was used and RNase T1 was added during the last 20 min of transcription (lanes 5 and 11), no change in the +1 initiated signals was observed as compared to lanes 3 and 4 or lanes 9 and 10, respectively.

The arrow labeled endogenous U1 in Figure 2B indicates the position of endogenous U1 that is known to undergo end-labeling in nuclear extracts if labeled UTP is used (data not shown). RNase T1 digestion removed the endogenous U1 band, demonstrating that the nuclease was working in the reaction. Lane 6 shows that higher levels of 3'-O-methyl-GTP had no effect on the accumulation of the UIG-initiated signals was observed as compared to lanes 3 and 4 or lanes 9 and 10, respectively.

The recovery of added control RNA was similar in each sample (band labeled control RNA). The HTFRG- and UIG- promoters are all inhibited by levels of α-amanitin that inhibit RNA polymerase II transcription [1 µg/ml] but that do not affect the RNA polymerase III-transcribed 5S maxigene. Additionally, the UIG-, HTFRG-, and AdMLG- promoters showed the same inhibition profile to increasing levels of α-amanitin, between 0 and 1 µg/ml, suggesting that they use the same RNA polymerase [data not shown]. No inhibition was observed if α-amanitin was added at the end of the transcription reaction, demonstrating that it had no RNase activity. Tagetitoxin is a recently characterized inhibitor of RNA polymerase III-directed transcription (Steinberg et al. 1990). The 5S maxigene was highly sensitive to tagetitoxin, whereas the UIG-, HTFRG-, and AdMLG- promoters were relatively unaffected.

Identification of the RNA polymerase synthesizing the UIG- and HTFRG- transcripts

Figure 3 tests the α-amanitin and tagetitoxin sensitivity of the UIG- and HTFRG- transcription signals as compared to the adenovirus major late (AdMLG-) and 5S maxigene promoters (Sakonju et al. 1980; Sawadogo and Roeder 1985). The recovery of added control RNA was similar in each sample [band labeled control RNA]. The HTFRG-, UIG-, and AdMLG- promoters are all inhibited by levels of α-amanitin that inhibit RNA polymerase II transcription [1 µg/ml] but that do not affect the RNA polymerase III-transcribed 5S maxigene. Additionally, the UIG-, HTFRG-, and AdMLG- promoters showed the same inhibition profile to increasing levels of α-amanitin, between 0 and 1 µg/ml, suggesting that they use the same RNA polymerase [data not shown]. No inhibition was observed if α-amanitin was added at the end of the transcription reaction, demonstrating that it had no RNase activity. Tagetitoxin is a recently characterized inhibitor of RNA polymerase III-directed transcription (Steinberg et al. 1990). The 5S maxigene was highly sensitive to tagetitoxin, whereas the UIG-, HTFRG-, and AdMLG- promoters were relatively unaffected.

Analysis of the 5' ends of UIG- and HTFRG- transcripts by primer extension

Figure 4 shows a direct analysis of the 5' ends of transcripts produced by both the UIG- [pSG611] and HTFRG- [pSG628] promoters. The end-labeled oligonucleotide used for primer extension is complementary to sequences in the UIG- and HTFRG- templates from +59 to +41 relative to the initiation site. Dideoxy sequencing of the UIG- promoter with the same labeled oligonucleotide generated a convenient DNA ladder. Note that the dideoxy sequencing DNA ladder for HTFRG- is not included because it is identical to UIG- in this region from -10 to +189. Transcription conditions were the same as those used to generate lanes 5 and 11 in Figure 2B. As can be seen in Figure 4, lane 5, the UIG- promoter had two major primer extension bands corresponding to initiation at -2 and +1. In Figure 2B, lane 5, the upper band of the doublet is the -2 start and the lower band is the +1 start. These identifications were confirmed by excising and recovering the RNA from the two bands in Figure 2B, lane 5, and mapping their 5' ends separately by primer extension [data not shown]. As can be seen in Figure 4, lane 6, primer extension analysis of the HTFRG- transcript gave rise to only one band corresponding to initiation at +1.

Analysis of the 3' ends of UIG- and HTFRG- transcripts by RNase U2 protection

It is interesting that the HTFRG- and UIG- +1-initiated transcripts differ in their initiation by several nu-
Correct transcription from human U1 promoter

cleotides [Fig. 2B, cf. lanes 5 and 7]. This implies that the U1G-+1-initiated transcript (and also the −2-initiated RNA) is several nucleotides shorter at the 3' end. This is shown directly in Figure 5. As diagramed in Figure 5A, a DNA oligonucleotide complementary to the 3' end of the G- cassette was annealed to RNA derived from either HTFRG- or U1G-. Next, the RNA-DNA hybrid was digested with RNase U2. RNase U2 cuts RNA with a high degree of specificity at ApN but does not cleave RNA that is duplexed to DNA. After digestion with RNase U2, the protected RNAs were analyzed by denaturing polyacrylamide gel electrophoresis (PAGE).

Figure 5B shows the results of RNase U2 analysis of the 3' ends of U1G- and HTFRG- transcripts. Lanes 1 and 3 are an analysis of HTFRG- and U1G- RNA, respectively, made under conditions favorable for +1 initiation. RNA with a 3' end at +189 should result in a 34-nucleotide protected band, whereas RNA with a 3' end at +186 should result in a 31-nucleotide band. The positions of these bands are indicated at the right of Figure 5B. To assist in this analysis, RNA size markers were included as described below.

Figure 5B, lane 2, is an analysis of U1G- RNA made under low (0.1 mM) 3'-0-methyl-GTP conditions favorable for the formation of readthrough transcripts. These conditions are identical to those found in Figure 2B, lane 3, and result in 20-fold more accumulation of readthrough transcripts than +1 transcripts (cf. readthrough to +1 transcripts in Fig. 2B, lane 3). Readthrough transcripts are expected to have 3' ends at positions +189 and +191, which should give rise to protected RNAs of 34 and 36 nucleotides, respectively. RNA with 3' ends longer than +191 are not protected since the protecting oligonucleotide extends only to +191. As expected, the 34-nucleotide band in Figure 5B, lane 2, comigrated with the HTFRG- band in lane 1, which demonstrates further that +1-initiated HTFRG- transcripts have a 3' end at +189.

In addition to cleaving RNA at ApN, RNase U2 also cleaves RNA at C, G, and U residues, although at a much lower efficiency. This nonspecific cleavage generates a faint RNA ladder. Extra sample was loaded in

Figure 4. Primer extension analysis of the 5' end of UIG- and HTFRG- transcripts. An oligonucleotide containing sequences complementary to the UIG- cassette from +59 to +41 was used for primer extension analysis of both the UIG- and HTFRG- transcripts [lanes 5 and 6, respectively] and to generate a marker DNA ladder by dideoxy sequencing of the UIG-template [lanes 1-4]. [Left] The DNA sequence for the UIG-template. The HTFRG- sequence is not shown but is identical to U1 from −9 to +191 (see Fig. 1). The nucleotide at position +1 is indicated by an arrow.

Figure 5. 3'-End analysis of UIG- and HTFRG- transcripts using an RNase U2 protection assay. [A] Schematic of the procedure for 3'-end analysis. The sequences of the 3' end of the UIG- RNA and HTFRG- RNA are shown along with the complementary DNA oligonucleotide. The numbers above the sequence denote the positions relative to the +1 start site. RNase U2 cleaves specifically at ApN but does not cleave RNA that is duplexed to DNA. After digestion with RNase U2, the protected RNAs were analyzed by denaturing polyacrylamide gel electrophoresis (PAGE). For lanes 1–3, standard transcription reaction conditions were used, except 100 μCi of [α-32P]CTP was used in place of labeled UTP and unlabeled UTP (0.4 mM) in place of unlabeled CTP. Transcription products were gel-purified and underwent the same annealing and RNase U2 digestion steps as described in Materials and methods. [Lanes 1 and 3] Transcription of the HTFRG- templates, respectively, under conditions favorable for +1 initiation (see Fig. 2, lanes 5 and 11). [Lane 2] Transcription of UIG- under low 3'-0-methyl-GTP conditions (0.1 mM) that favor the synthesis of readthrough transcripts (same conditions as in Fig. 2, lane 3). [Lane M] DNA marker lane as described in Fig. 3. (Right) positions and 3' ends of the protected RNAs. The position of the 26-nucleotide size marker is also indicated.
Figure 5B, lane 2, to permit this RNA ladder to be visible. Using this ladder, the difference between the bands in Figure 5B, lanes 1 and 3, is determined to be 3 nucleotides. This indicates that the UIG" +1 transcript has a 3' end at +186. Note that a 26-nucleotide RNA was added to each sample to ensure that the RNA products in each lane migrated at the same rate. Much less of this 26-nucleotide size marker was added to lane 3 because the UIG" +1 transcript is considerably less abundant than either the readthrough or HTFRG" transcripts [Fig. 2].

One other observation supports the evidence that the 3' ends of UIG" transcripts correspond to the 3' end found in vivo for the wild-type U1 gene. In the UIG" template, position +186 is 5 bp upstream of the 3'-end box [see Fig. 1]. When the wild-type UIG" gene is injected into oocytes, the 3' ends of the resulting U1 RNAs are found at position +167 as well as positions +168 and +169. Position +167 is also 5 bp upstream of the 3'-end box in the wild-type gene.

Effects of promoter mutations to the UIG" construct

Although we have demonstrated that the UIG" construct initiates and terminates where one would expect, it could still be argued that the -2-186-terminated and +186-terminated transcripts are artifacts resulting from the insertion of the G'-cassette and addition of 3'-O-methyl-GTP into the reaction. To establish further the validity and to demonstrate the utility of this system, the effects of mutations in known snRNA promoter elements on transcription in vitro were assessed. Figure 6 demonstrates that +1 transcriptional activity responds to promoter mutations known to have in vivo effects. Deletion of the DSE [UIG" -DSE] in construct pSG626 [lane 6] results in a sixfold drop in +1 transcription activity. The level of readthrough transcription is unaffected as would be expected if it were due to mRNA-type transcription.

Deletion of the PSE [UIG" -PSE], as is done in construct pSG627 [Fig. 6, lane 5], results in a threefold drop in +1 transcription activity. One might expect a much larger effect in deleting the PSE. In this deletion construct, however, a sequence formerly at -110 that closely resembles the U1 PSE is moved into the PSE position. This sequence is GTGTCAGGG, as compared to the wild-type sequence GTGACCCTG, and is shifted 1 bp closer to the initiation site. It may be replacing the normal PSE functionally, which would explain the small effect of the PSE deletion. To clarify the issue, we replaced the PSE region with an oligonucleotide that has no obvious homology to the PSE and does not change the spacing between the other U1 promoter elements. As shown in Figure 6, lane 3, the resulting template, UIG" -repPSE [replace the PSE], has almost no +1 transcription activity [20-fold decrease].

Larger scale deletions such as those found in the UIG" construct [Fig. 6, lane 6] and in the G'-cassette construct [not shown] also abolish +1 transcription activity [>20-fold decrease]. Because the signal is relatively faint, decreases >20-fold cannot be estimated accurately. Thus, the effects of promoter mutations may be much greater than 20-fold. Intriguingly, deletion of the 3'-end box [UIG" -Δ3' box] decreased +1 transcription activity by 10-fold [lane 4], establishing this region as an important promoter element for this system. Finally, when U1 sequences from -37 to -6 were replaced with an unrelated oligonucleotide lacking homology to TATA or PSE elements [UIG" -rep -37/-6], full-strength +1 transcription activity was observed [lane 7]. The oligonucleotide replaced a poor TATA homology [TATGAG] found at -33 of U1, thus showing it to not be a functional element in this system. The small decrease seen between lanes 1 and 7 is not significant since the control RNA decreased proportionately.

U1G" transcription is sensitive to PSE1, a transcriptional activating protein

As reported previously, immunoprecipitation of crude nuclear extracts with monoclonal antibodies raised against the smaller subunit of the heterodimeric protein PSE1 decreased transcriptional activity of a promoter consisting of the minimal HTFR promoter fused to the U1-coding region. This effect was reversed by the addition of purified PSE1 protein, while transcriptional activity of the AdML promoter was unaffected by either depletion or repletion of PSE1 [Knuth et al. 1990]. Figure 7, lanes 1-3, indicates that the transcription of an HTFRG" promoter was also sensitive to levels of PSE1, and lanes 7-9 show that transcriptional activity of an
reaction parameters we were able to improve the signal was extremely weak; however, by varying transcription known Ul promoter elements, sequences at the Ul initiation site, and the 3'-end box. Additionally, a transcription activating protein, PSE1, is inserted immediately downstream of a human Ul promoter retaining all of the competent to drive transcription of a G\(^{-}\) cassette (Sawado and Roeder 1985) which will allow for the general study of vertebrate snRNA transcription in completely homologous systems. HeLa cell nuclear extracts, which will allow for the qualitative recovery of RNAs from each sample. The control RNA for lanes 1–6 are not shown but varied less than twofold.

AdMLG\(^{-}\) promoter remained insensitive to this level of depletion or repletion. These results demonstrate that the previous work on PSE1 can be extended to the G\(^{-}\) cassette assay and serve as positive and negative controls for the examination of effects on U1G\(^{-}\) transcription. As can be seen by comparing lanes 4 and 5, specific immunodepletion of nuclear extract for PSE1 resulted in a 20-fold loss of signal from the −2 and +1 starts in the U1G\(^{-}\) system. Lane 6 indicates that when purified PSE1 (>95% pure) was added, transcription from the −2 and +1 starts was restored by 10-fold. Addition of purified PSE1 protein did not detectably stimulate transcriptional activity of the PSE replacement construct (repPSE).

Discussion

In this report the development of an in vitro transcription system for the human U1 snRNA gene is described. Additionally, a transcription activating protein, PSE1, is shown to be involved in the activity of this system. This is the first in vitro transcription system for RNA polymerase II-transcribed vertebrate snRNA genes that uses HeLa cell nuclear extracts, which will allow for the general study of vertebrate snRNA transcription in completely homologous systems. HeLa cell nuclear extracts prepared by the method of Dignam et al. (1983) were competent to drive transcription of a G\(^{-}\) cassette (Sawadogo and Roeder 1985) inserted immediately downstream of a human U1 promoter retaining all of the known U1 promoter elements, sequences at the U1 initiation site, and the 3' end box.

The initial transcription signal we observed for U1G\(^{-}\) was extremely weak; however, by varying transcription reaction parameters we were able to improve the signal by 10-fold (see Fig. 8). In addition to other parameters, optimal +1 transcription activity occurs in a narrow window centered around 3.5 mM MgCl\(_2\). When the wild-type U1 gene was tested under the conditions optimal for U1G\(^{-}\), no detectable +1 initiation activity was
found, even when MgCl$_2$ was varied in small increments.

We show that +1-initiated transcripts are driven by RNA polymerase II and efficiently utilize the snRNA-specific 3'-end box to direct 3'-end formation. When the wild-type U1 gene is injected into $X$. laevis oocytes, the 3' end of the cytoplasmic UI precursor occurs at positions +167, +168, and +169 [Neuman de Vegvar et al. 1986]. As can be seen in Figure 1, position +186 of the UIG$^-$ template is equivalent to +167 in the wild-type U1 if one measures 5 bp upstream from the U1 3'-end box. Thus, both the 5' and 3' ends that are synthesized in vitro correspond to what is observed in vivo.

In addition to the properties mentioned above, this in vitro system has several other features consistent with the known properties of snRNA promoters. First, transcriptional activity is dependent on the upstream promoter elements PSE and DSE. Deletion of the DSE has less effect on in vitro transcription than it does in vivo [6-fold vs. 100-fold]. This difference may be due to a cryptic enhancer element being brought into proximity by the deletion. Additionally, mRNA enhancers almost always exhibit only a two- to threefold effect in vitro [for discussion, see Short 1988], so it is not necessarily surprising that we observe a sixfold effect. Replacement of the PSE with a sequence with no homology to the PSE [U1G$^-$ rep-PSE] resulted in a >20-fold reduction in transcriptional activity. A double point mutation in the highly conserved core of the PSE (changing the −58/−57 CC to TT) also results in a >20-fold decrease in transcriptional activity. Thus, the U1G$^-$ system is highly sensitive to snRNA-specific mutagenesis. Finally, it could be argued that the +1 start is driven by a cryptic TATA box sequence found near −30. However, in the template UIG$^-$ rep−37/−6, where this region is replaced with a sequence with no homology to TATA or PSE-like elements, +1 transcriptional activity is unaffected.

A second feature of this system is that transcription initiation at +1 is dependent on the presence of the 3'-end box. Deletion of the 3'-end box reduces transcription initiation at +1 by 10-fold. Furthermore, deletion of upstream promoter elements fails to uncouple initiation from 3'-end formation. For example, for all of the mutations in UIG$^-$ that are shown in Figure 1, transcripts initiating at +1 form their 3' ends at position +186, not at +189.

Transcription initiation in vivo of snRNA genes occurs precisely and only at position +1 [unless heterologous systems are used]. The presence of the −2 start signal is the only feature of our system that is not similar to snRNA transcription. The fact that the −2 start terminates where the +1 start begins is active as the +1 start signal, and is responsive to the same promoter elements and to PSE1 protein suggests that it has many of the properties of an snRNA-type signal. Why position −2 is chosen as an initiation site is unclear but may reflect the use of an overlapping PSE element or that the snRNA transcription complex and its components have been altered slightly during the preparation of nuclear extracts. Heterogenous starts are also observed (even in homologous systems) when mutant snRNA genes are used [Hernandez and Weiner 1986; Neuman de Vegvar et al. 1986; Ciliberto et al. 1987; Dahlberg and Schenborn 1988].

U1-specific transcripts can be visualized only when enough 3'-O-methyl-GTP and RNase T1 are present to inhibit transcription arising from other parts of the U1G$^-$ template [see Fig. 2]. The +1-initiated transcripts are not detectable by other sensitive transcription assays unless 3'-O-methyl-GTP is present [data not shown]. This observation may explain why in vitro systems for vertebrate snRNA genes have been difficult to develop, namely that snRNA synthesis is readily inhibited by transcription complexes (presumably mRNA-like) initiating upstream of the U1 initiation site. Inhibition could be due to the elongating complex disrupting the formation of U1-specific initiation complexes or dislodging already formed U1-specific initiation complexes. Either case is compatible with the concept that U1 transcription, and perhaps snRNA transcription in general, is compartmentalized in vivo. It has been proposed that compartmentalization prevents the "squelching" effect that would occur if highly active genes were transcribing in proximity with genes that are lower in activity [for a recent discussion, see Tanaka et al. 1988]. In an alternative hypothesis, loss of compartmentalization may allow mRNA transcription complexes or transcription factors to inhibit snRNA promoters. The function in vivo of this inhibition may be to suppress transcription of a snRNA gene that happened to be in proximity of mRNA genes. This would explain the long-standing paradox of why snRNA genes (except for U6), which are highly active in vivo, are not active in most in vitro systems.

The results presented in this paper address the issue of whether 3'-end formation is the result of processing a longer, short-lived precursor or the result of terminating the snRNA-specific transcription complex at that site [for a recent discussion, see Neuman de Vegvar and Dahlberg 1989]. Our results suggest the synthesis and processing of a longer precursor in vitro to be unlikely, because the longest precursor can only be 3 nucleotides longer than the final transcript given that 3'-O-methyl-GTP efficiently chain-terminates all transcripts at +189. Further mutagenesis in the 3'-end box is also being done to determine how this element functions.

The only other in vitro transcription system for vertebrate snRNAs uses homogenates of manually isolated X. laevis oocyte GVls to direct snRNA synthesis [Lund and Dahlberg 1989] and has certain weaknesses and advantages relative to the UIG$^-$ system. One advantage is that the GV system has been shown to transcribe faithfully a variety of snRNA genes from several species and can use the wild-type gene. We have not yet examined whether our results in the UIG$^-$ system can be reproduced using other snRNA promoters, but we note that the human UIG$^-$ template is transcribed faithfully in mouse cell nuclear extracts [data not shown] and in GV homogenates [E. Lund, pers. comm.].

In the GV system, the homogenate is prepared with
minimal disruption and handling of the nuclei. Since isolated mammalian cell nuclei only retain snRNA transcriptional activity if the nuclei are isolated in isotonic, nonionic media such as sucrose [Lobo and Marzluff 1987], leakage or loss of factors from nuclear extract preparations is a real possibility. The transcriptional activity of the U1G− system is low relative to the GV system, which limits the range of effects that can be observed by mutagenesis of the promoter or by depletion of factors.

The major strength of the U1G− system relates to its utility for detection, fractionation, purification, and characterization of transcriptional activators. Preparation of large quantities of nuclear extracts for the U1G− system is easier than for the GV system, which is more suitable for analytical work. Also, the availability of oocytes is a fairly unique feature of Xenopus, which leaves the researchers to study their particular snRNA gene in a heterologous system. Since much more is known about human transcription factors than their Xenopus counterparts, use of the G− cassette will also allow the researcher to more easily examine the interactions of snRNA-specific transcription factors with other more characterized components of the transcription machinery.

The utility of this system is underscored by the identification of PSE1 as a potential transcription factor for U1. Verification that PSE1 is a bona fide U1 transcription factor will have to await in vivo tests. Previously, we reported that purified SP1 and major late transcription factor (MLTF) are able to footprint to sequences within the human U1 promoter elements [Gunderson et al. 1988]. Other transcription factors have also been implicated, including members of the octamer-binding transcription factor (OTF) family [Dahlberg and Lund 1988]. These factors have been implicated either on the basis of binding data or by observing the effects of promoter mutations on transcriptional activity. We are now in a position to directly analyze, by depletion/repletion experiments, the role of these transcription factors in the U1G− transcription system.

The reported similarity of PSE1 to the lupus autoantigen Ku is intriguing in several ways. First, Ku has been shown by immunolocalization to reside on subnuclear structures [Reeves 1987], consistent with the hypothesis that transcription of snRNAs might be localized within the nucleus. Second, the fact that Ku is a member of the set of lupus autoantigens, which includes snRNP proteins, and the La protein, involved in transcription termination of small RNAs, is also supportive of its role in this gene family. Further studies of the functional domains and DNA-binding modes PSE1/KU may illuminate its role in snRNA transcription activation.

In the previous paper discussing PSE1 [Knuth et al. 1990], the use of linear templates in the transcription reactions left open the possibility that PSE1 activated transcription by suppressing nonproductive end-to-end transcription. The results presented in Figure 7 are done with a circular template, making this interpretation of the data unlikely.

Materials and methods

Enzymes and reagents

DNA-modifying enzymes, AMV reverse transcriptase, and RNase inhibitor (RNasin) were obtained from Promega or New England Biolabs and used as per manufacturers’ instructions. NTPs labeled with 32P were obtained from DuPont New England Nuclear. RNase T1, RNase U2, and 3′-O-methyl-GTP were obtained from Pharmacia. DNA oligonucleotides were synthesized by the University of Wisconsin Biotechnology Center and were purified by gel electrophoresis prior to use. α-Amanitin was from Sigma, and tagetitoxin (Tagetin), which is an inhibitor of RNA polymerase III transcription [Steinberg et al. 1990], was from Epicentre Technologies [Madison, WI].

Nuclear and whole-cell extracts

HeLa S3 cells were grown in either Joklik’s or S-MEM media supplemented with 5% calf serum, nonessential amino acids, glutamine, and antibiotics (GIBCO). Cells were harvested at 4 × 106 to 9 × 106 to cells/ml, and nuclear extracts were prepared according to the method of Dignam et al. [1983], with 40 μM Tris-HCl (pH 7.9) substituted for HEPES in the buffers. S-100 whole-cell extracts were prepared by the method of Weil et al. [1979]. The optional steps of dialyzing the S-100 extract or supplementing it with 50% glycerol were not done.

Construction of DNA templates

The U1 gene sequences from −393 to +199 are originally from pHU1-1D [Lund and Dahlberg 1984] and were inserted into the BamHI site of pTZ18U. To facilitate the cloning steps, a unique SnaBI site was put into the U1 gene at +8 by changing position +9 C to G and +11 G to A, and a unique StuI site was put in at −38 by changing position −36 G to C. Because these point mutants fall outside of the U1 promoter elements, they did not affect promoter activity as assayed in oocyte injections [data not shown]. Construct pSG644 (see Fig. 1) was made first, by replacing the U1 coding sequences downstream of +9 by blunt-ligating 179 bp of the G− cassette [from +10 to +188 in Sawadogo and Roeder 1985] to SnB1- and HinClI-restricted U1 DNA. SnB1 cuts at +8, and HinClI cuts just downstream of +199. pSG611 was made from pSG644 by ligating a double-stranded oligonucleotide containing the U1 3′-end box sequence from +172 to +191 to AYCl1- and AYCl2- and AvrII-digested pSG644. A schematic of pSG611 and pSG644 is found in Figure 1, and both were used to make the mutated U1G− promoters. All plasmids were verified by restriction analysis and Sanger dyeoxy and Maxam–Gilbert sequencing [Sanger et al. 1977; Maxam and Gilbert 1980]. Details of mutated constructs are found in Figure 1. pSG638 replaces the PSE region of U1G− from −77 to −38 with an oligonucleotide of the same length with the sequence (top strand) 5'-GAGAGTACGAGTGGAAATGGAGCCCT-GCACCGAGAGAG-3'. pSG636 replaces U1G− sequences from −37 to −6 with an oligonucleotide of identical length whose sequence (top strand) is 5'-CCTCATCTCTCTCTCT-TAAG-3'. pSG628 was made by replacing U1G− sequences from −393 to −8 with sequences from the HTFR promoter from −100 to −8. This HTFR sequence corresponds to the minimal HTFR promoter that still retains full activity in vivo [Casey et al. 1988]. The HTFR− promoter was highly active in our extracts, initiating at the U1

GENES & DEVELOPMENT 2057
+1 site and was a useful size marker in our analysis of the U1G− promoter.

In vitro transcription

Unless stated otherwise, the in vitro transcription reactions for the U1G− and HTFRG− templates were done in 25 µl, using conditions determined to be optimal for +1 transcription activity (see Figs. 2 and 8), which are 1.8 µg of circular plasmid DNA [0.7 pmol], 9 µl of nuclear extract (8 mg protein/ml), along with 18 mM Tris-HCl [pH 7.4 at 25°C], 44 mM KCl, 0.05 mM EDTA, 9% glycerol, 0.2 mM dithiothreitol, 0.4 mM ATP, 0.4 mM CTP, 0.8 mM 3'-O-methyl-GTP, 0.4 units RNase T1 throughout the reaction, 0.26 µM [α-32P]UTP [20 µCi], and 3.5 mM MgCl2. Identical reaction conditions were used for the transcription of the AdML promoter, except 6 mM MgCl2 was used. The AdML promoter was inserted upstream of a 375-bp G− cassette (designated here as AdMLC−) and was kindly provided by M. Sawadogo and R. Roeder [1985]. All ingredients were mixed prior to addition of the extract, and transcription was allowed to proceed for 90 min at 30°C. The RNA products were extracted by the addition of 150 µl TESS [0.1 mM Tris-HCl at pH 7.9, 2 mM EDTA, 0.5% SDS, 0.3 M sodium acetate, and 3 µg/ml tRNA] and 300 µl buffered phenol. The phenol was back-extracted with 150 µl of TESS, and the aqueous phases were pooled, ethanol-precipitated, washed, dried in vacuo, and resuspended in formamide loading buffer. To follow the quantitative recovery of RNA during extraction, precipitation, and gel loading, labeled RNA was added to the TESS during extraction. The labeled RNA is a 432-nucleotide transcript synthesized by T7 RNA polymerase and is designated control RNA in the figures. PAGE used 7% (30 : 0.8) acrylamide gels with 8 M urea. Autoradiography was done using an intensifying screen where indicated. The transcription activity was determined by measuring the absorbances of the autoradiographic bands using a Biomed Instruments scanning densitometer (Biomed Instruments, Chicago, IL).

Transcription of the Xenopus borealis somatic SS RNA maxigene [kindly provided by Donald Brown [Sakonju et al. 1980] was done using 10 µl of S-100 extract [undilayed] and 1.5 µg of circular plasmid DNA. Net solution conditions were as follows: 20 mM Tris-HCl [pH 7.4 at 25°C], 60 mM KCl, 2 mM dithiothreitol, 0.4 mM AMP, 0.4 mM CTP, 0.4 mM UTP, 0.01 mM GTP, 0.13 µM [α-32P]GTP [10 µCi], and 1.8 mM MgCl2. Transcription proceeded for 1 hr at 30°C, and the RNA products were extracted and analyzed as described above.

Analysis of 5' and 3' ends of RNA transcripts

An oligonucleotide complementary to +59 to +41 of the U1G− template was end-labeled for primer extension using polynucleotide kinase and γ-32P[ATP. For primer extension, the transcription reactions were done in the absence of labeled UTP, and the RNA products were extracted by a method employing 4 M guanidinium-HCl and acidic phenol (Chomczynski and Sacchi 1987). After precipitation, the RNA and labeled primer were annealed in 15 µl of 50% formamide, 0.4 M NaCl, and 0.2 M PIPES [pH 6.6] for 6 hr at 33°C. The RNA–DNA duplex was then extended, using 3 units AMV reverse transcriptase and 80 units RNase inhibitor for 30 min at 42°C. The products were extracted and analyzed on 7% polyacrylamide gels with 8 M urea.

For 3'-end analysis, the transcription reactions were done with [α-32P]CTP [100 µCi] in place of labeled UTP as described in Figure 5. RNA products were extracted with TESS, as described above, and separated by denaturing PAGE. RNA transcripts were excised and eluted from the gel and annealed to a DNA oligonucleotide complementary to the 3' end of the RNA whose sequence is given in Figure 5A. The RNA–DNA duplex was then digested with 20 units of RNase U2 for 20 min at 37°C, and the products were extracted and analyzed by denaturing PAGE.

Depletion/repletion of nuclear extracts for PSE1

Nuclear extracts were specifically depleted for PSE1 protein as described previously [Knuth et al. 1990]. Briefly, a monoclonal antibody recognizing the smaller subunit of PSE1 was bound to fixed Staphylococcus aureus cells. Small amounts of the cells (<2% of extract volume) were added to the extracts and allowed to bind PSE1 for 1 hr on ice. The cells were pelleted, and the supernatants were tested for transcription on the HTFRG− template. An optimal level of cells was used for testing the U1G− and AdMLG− promoters that significantly decreased HTFRG− transcription, yet allowed significant repletion of activity upon addition of PSE1. Depleted samples were assayed under standard methods described; and for the repleted extracts, 1 µl [50 ng] of PSE1 purified from HeLa cells as described [Knuth et al. 1990] was added to 9 µl of extract just before starting the reactions. Neither mock depletion of extracts with cells alone nor addition of 1 µl of the PSE1 storage
buffer to PSE1-depleted extracts had any affect on transcription of the three templates (data not shown).

Acknowledgments

We thank James Dahlberg, Peggy Farnham, and Janet Mertz for critically reviewing this manuscript. We thank Nancy Thompson for her help in preparing the PSE1 monoclonal antibodies, Elsebet Lund for helpful guidance and oocyte injections, Lee Strasheim for technical help, and members of the Burgess lab for helpful comments. We also wish to acknowledge Donald Brown for his gift of the 5S maxigene and Robert Roeder and Michelle Sawadogo for their gift of the G-less cassette and the adenovirus 2 major late G-less cassette.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References


Neuman de Vegvar, H.E., E. Lund, and J.E. Dahlberg. 1986. 3′ End formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* 47:259–266.


Gunderson et al.


The human U1 snRNA promoter correctly initiates transcription in vitro and is activated by PSE1.

S I Gunderson, M W Knuth and R R Burgess

*Genes Dev.* 1990, 4:
Access the most recent version at doi:10.1101/gad.4.12a.2048

References

This article cites 36 articles, 14 of which can be accessed free at: http://genesdev.cshlp.org/content/4/12a/2048.full.html#ref-list-1

License

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.