Nucleic Acids Research, 1996, Vol. 24, No. 22 **4525–4534**

Increasing the distance between the snRNA promoter and the 3′ box decreases the efficiency of snRNA 3′-end formation

Lakshman Ramamurthy[1,3], Thomas C. Ingledue[2], Duane R. Pilch[1], Brian K. Kay[2,3] and William F. Marzluff[1,2,3,*]

[1]Program in Molecular Biology and Biotechnology, [2]Curriculum in Genetics and Molecular Biology and [3]Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA

Received July 22, 1996; Revised and Accepted September 25, 1996

ABSTRACT

Chimeric genes which contained the mouse U1b snRNA promoter, portions of the histone H2a or globin coding regions and the U1b 3′-end followed by a histone 3′-end were constructed. The distance between the U1 promoter and the U1 3′ box was varied between 146 and 670 nt. The chimeric genes were introduced into CHO cells by stable transfection or into *Xenopus* oocytes by microinjection. The efficiency of utilization of the U1 3′ box, as measured by the relative amounts of transcripts that ended at the U1 3′ box and the histone 3′-end, was dependent on the distance between the promoter and 3′-end box. U1 3′-ends were formed with >90% efficiency on transcripts shorter than 200 nt, with 50–70% efficiency on transcripts of 280–400 nt and with only 10–20% efficiency on transcripts >500 nt. Essentially identical results were obtained after stable transfection of CHO cells or after injecting the genes into *Xenopus* oocytes. The distance between the U1 promoter and the U1 3′ box must be <280 nt for efficient transcription termination at the U1 3′ box, regardless of the sequence transcribed.

INTRODUCTION

A novel feature of the biosynthesis of vertebrate snRNAs is that transcription must initiate at an snRNA promoter for 3′-end formation (1,2). There is a single required sequence element for 3′-end formation, the 3′ box, located ~10 nt downstream of the end of the primary transcript (3) and there is no requirement for sequences in the mature snRNA for proper 3′-end formation (1). However, when a heterologous promoter [e.g. thymidine kinase (2), globin, adenovirus (1) or histone (4)] is used in place of the snRNA promoter, the snRNA 3′-end is not formed; rather, transcription continues past the normal 3′-end and the transcripts are polyadenylated using cryptic or natural polyadenylation sites (1,2). In addition, longer read-through transcripts formed in isolated nuclei are not precursors of mature U1 RNA molecules (5) and longer 'precursors' are not converted to mature snRNAs when they are injected into *Xenopus* oocytes (6). These results suggest that a sequence in the snRNA promoter is required for proper 3′-end formation and that 3′-end formation occurs co-transcriptionally, either as a termination event or as a very rapid processing event. Following formation of the primary transcript, the mature snRNA is formed by removing nucleotides (<15) from the 3′-end, presumably by an exonuclease(s) in the cytoplasm (5,7), while the removal of the last 2 nt takes place in the nucleus (8).

We report here that there is a strong distance dependence for the coupling of initiation from vertebrate snRNA promoters with formation of snRNA 3′-ends. By varying the amounts of histone or α-globin sequence between the promoter and the 3′ box, we have constructed genes that encode transcripts ranging in length from 146 to 670 nt, ending at the snRNA 3′-end. We find that the snRNA end is formed inefficiently, as detected by the preferential formation of the distal histone 3′-end, if the transcript is longer than 500 nt. If the distance between the transcription start site and the U1 signal is <200 nt, then the U1 3′-ends are formed very efficiently. Similar results were obtained both in mammalian cells and in *Xenopus* oocytes.

MATERIALS AND METHODS

Construction of cloned genes

The chimeric U1 and histone H2a genes were constructed from the mouse histone H2a-614 gene (9,10) and the mouse U1b.1 and U1b.2 genes (11,12); these genes are shown in Figure 1. The mouse U1b promoter containing 5 nt of U1 coding sequence has been described previously, as have the cassettes containing the U1 3′ signal with either 10 or 50 nt of U1 coding sequence (4) and the histone H2a-614 3′-end signal (13). The genes are named by the length of the snRNA transcript they produce and whether they have 49 (UL genes) or 10 nt (US genes) of U1 RNA sequence at their 3′-end.

There is a translation initiation codon within 60 nt of the start of transcription in all the genes, as well as an in-frame translation termination codon prior to the U1 3′-end, except for the UL_{190} gene and US_{146} genes, which have a stop codon just after the U1 3′-end. We constructed another set of genes that contain human α-globin cDNA sequences in place of the histone sequences. These clones were constructed using UL_{172} as the parent clone to which was attached varying portions of human α-globin cDNA sequence. The entire α-globin cDNA sequence was included in the UG_{670} gene

*To whom correspondence should be addressed. Tel: +1 919 962 8920; Fax: +1 919 966 6821; Email: marzluff@med.unc.edu

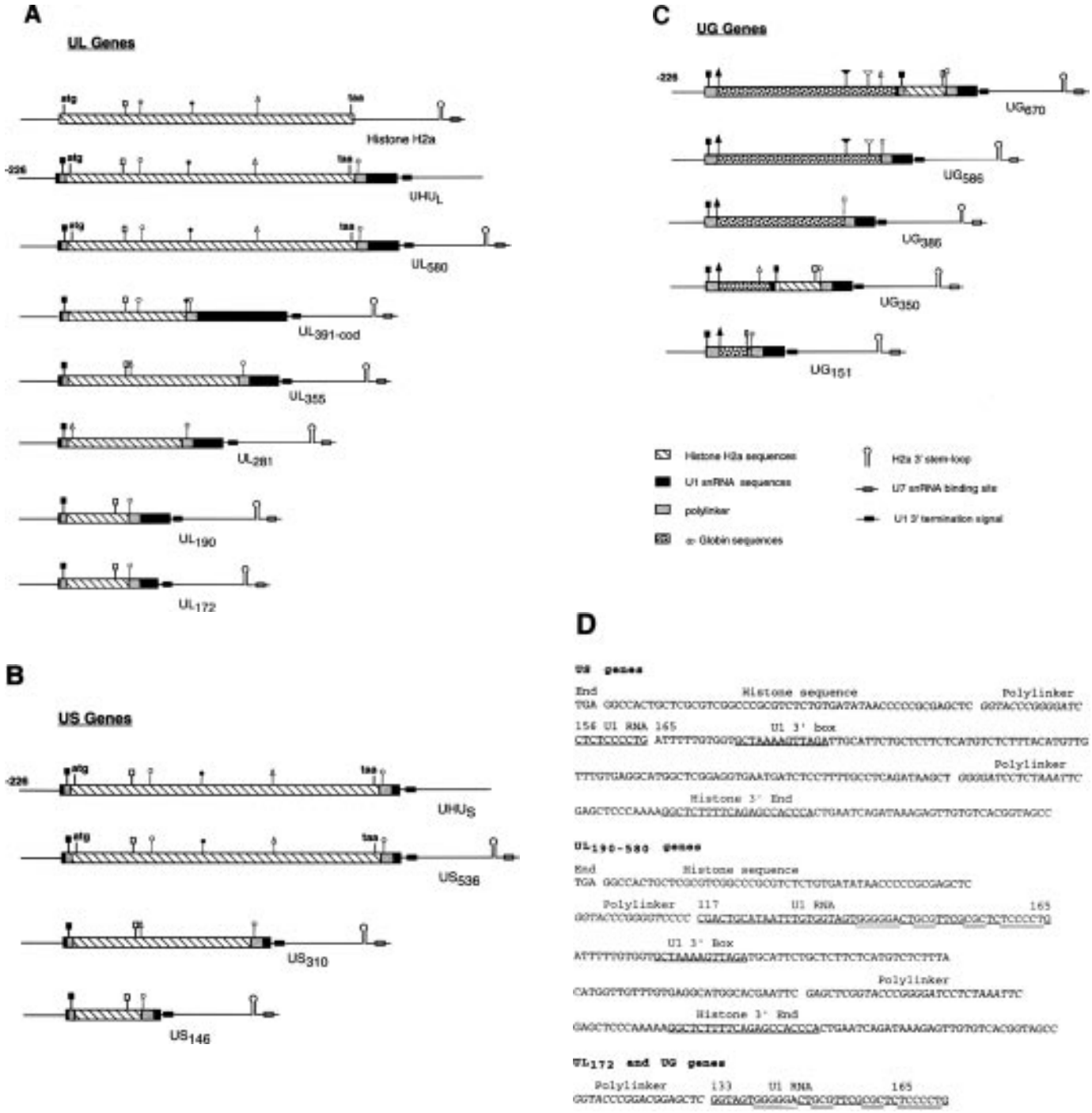

Figure 1. Structure of histone–U1 chimeric genes. The UL (**A**) and the US (**B**) genes. The mouse histone H2a-614 gene (pH2a) is shown at the top of (A). The UHU$_L$ and UHU$_S$ genes have been described previously (4). They contain the promoter and first 5 nt of the mouse U1b gene, the complete histone coding region and then the U1 3′-end starting at either nt 117 or 156 of the U1 coding sequence. The other genes contain the U1 promoter and first 5 nt of coding sequence, variable amounts of histone sequence and end with the last 49 (UL genes) or last 10 nt (US genes) of U1 snRNA followed by the histone H2a 3′-end. The UL$_{391\text{-cod}}$ gene has the complete U1 sequence attached at the *Nar*I site at codon 43 of the H2a gene. The U1 and histone coding sequences are indicated by the large boxes (see C). The restriction enzyme sites used in the S1 nuclease mapping experiments are shown. The restriction sites are: ∇, *Bst*EII; ▼, *Hin*dIII; Δ, *Msc*I; ●, *Nar*I; ▲, *Nco*I; ❍, *Sma*I; ❑, *Sty*I; ■, *Xba*I. (**C**) The structure of the UG genes is shown. The UG$_{386}$ and UG$_{586}$ genes have all the histone sequences removed and contain different amounts of globin cDNA followed by the UL$_{172}$ 3′-end. The UG$_{670}$ and UG$_{350}$ gene are identical to the UL$_{172}$ gene except for the insertion of human α-globin sequences at the 5′-end of the transcript. (**D**) The sequences at the junctions of the different 3′ processing signals are shown. The U1 coding sequences and 3′-end signal, polylinker and histone 3′-end signals are indicated. The polylinker sequences are in italics. The U1 3′ box signal, the U1 coding sequences and the stem–loop at the 3′-end of the histone mRNA are underlined. The stem–loop at the 3′-end of the U1 sequence in the U1 gene is double underlined. Note that the U1 snRNA gene which donated the U$_L$ 3′-end is the mouse U1.1 gene and the U1 gene which donated the U$_S$ 3′-end was from the mouse U1.2 gene (12). The portion of the UL$_{172}$ 3′-end which differs from the UL$_{190}$ 3′-end is shown. This sequence lacks the Sm binding site and is the 3′-end used in all the UG genes.

containing the translational start codon and a stop codon.The numbers refer to the length of the transcript that ends at the U1 3′-end. All of these UG clones contain the normal globin ATG codon and an in-frame stop codon prior to the U1 3′-end. The sequences of the 3′-ends of the genes with both the U1 and histone 3′-ends are shown in Figure 1D.

Transfection

The genes were introduced into CHO cells by co-transfection with the pSVneo gene using the polybrene procedure (15). Stable transfectants were isolated by selection with G418 as previously described (16). Pools of transfectants, 20–50 per flask, were pooled and grown in the absence of G418 for analysis of expression of the transfected genes.

Preparation and analysis of RNA

RNA was prepared from exponentially growing cells (<50% confluent) as previously described (16). The 3′-ends of the transcripts from the transfected genes were analyzed by S1 nuclease mapping using the probes described in the figure legends. The 5′-ends were mapped using probes labeled at an appropriate internal site in the gene.

The probe used in Figure 6D was made using PCR to amplify a 152 nt fragment of DNA containing the sequence from –13 to +127 of the U1–histone hybrid genes plus an additional 12 nt of non-homologous sequence included at the end of the 5′ primer. This fragment was labeled with polynucleotide kinase and [γ-^{32}P]ATP and used in an S1 nuclease protection assay.

The conditions of hybridization and digestion have been described (9). The protected fragments were resolved by electrophoresis on 6% polyacrylamide–7 M urea gels, detected by autoradiography and quantified by densitometry or on a PhosphorImager (Molecular Dynamics).

Injection of *Xenopus* oocytes

Supercoiled DNA (15 nl, 30 μg/ml) was injected into stage VI *Xenopus* oocytes (17) and the oocytes were incubated at 18°C for 18 h. In some experiments (noted in the figure legends) the amount of DNA injected was varied. RNA was prepared as previously described (18) and analyzed by S1 nuclease mapping as described above.

RESULTS

In the course of constructing genes which would express histone mRNAs ending in a U1 snRNA 3′-end, we observed that the U1 3′-end was formed inefficiently (14). To study some of the possible parameters affecting snRNA 3′-end formation in these chimeric genes, we constructed genes of varying lengths with a U1 promoter, either histone or globin coding sequences and a U1 3′-end followed by an efficient histone 3′ processing signal located ~100 nt 3′ of the U1 3′ box. Any transcripts which do not terminate at the U1 3′ box should be processed at the histone processing signal.

Figure 1 shows the genes used in these experiments. All the genes had a 226 nt mouse U1b promoter and the first 5 nt of the U1 RNA sequence. This cassette was fused to portions of either the mouse histone H2a-614 gene (UL and US genes; Fig. 1A and B) or the human α-globin coding region (UG genes; Fig. 1C). These coding regions were followed by a U1 3′-end, with either 49 (UL genes), 32 (UL_{172} and UG genes) or 10 nt (US genes) of U1 coding sequence followed by the U1 3′ box. The histone H2a-614 3′-end and processing signal were placed ~100 nt 3′ of the U1 3′ box. These genes were named according to the sequence in the gene, the amount of U1 coding sequence present and the length of the transcript ending at the U1 3′-end (indicated by the subscript). The UL genes contain the terminal stem–loop in the U1 RNA, while the US genes contain only the last 10 nt of U1 sequence and no secondary structures present in the U1 snRNA. The UL and US genes contain histone sequences while the UG genes contain globin sequences (Fig. 1C). For example, the UL_{580} gene encodes a 580 nt transcript ending with the last 49 nt of U1 snRNA and the US_{310} gene encodes a 310 nt transcript ending with the last 10 nt of U1 snRNA. Figure 1D shows the sequences at the 3′-ends of the genes.

In all these genes there are two functional 3′ signals such that there are two distinct transcripts produced from these genes, one ending at the U1 3′-end and the other at the histone 3′-end. The different 3′-ends were distinguished by S1 nuclease mapping and the distinct protected products resulting from transcripts ending at the histone or U1 3′-ends were quantified by densitometry or on a PhosphorImager. The relative amounts of transcripts with either U1 or histone 3′-ends gives a measure of the efficiency of U1 3′-end formation, assuming that the transcripts have similar stabilities. We assume that all of the transcripts which extend past the U1 3′-end are processed at the distal histone 3′-end. This is a good assumption, since the histone H2a-614 3′ processing signal is very efficient (13) and has been previously shown to be utilized efficiently on transcripts which initiate at the U1 promoter (4). The chimeric genes were introduced into CHO cells and pools of stable transformants assayed to determine the proportion of steady-state transcripts ending at the histone or U1 3′-ends. The genes were also injected into *Xenopus* oocytes to test expression in another cell type and to address the possibility that measurements of the relative amounts of transcripts with different 3′-ends in steady-state RNA in mammalian cells may not reflect the relative efficiency of 3′-end formation. Transcripts in *Xenopus* oocytes are generally stable and it has been possible to detect transcripts in oocytes which are often undetectable in somatic cells, such as the prematurely terminated *myc* (19,20) or α-tubulin transcripts (21).

snRNA 3′-ends are formed efficiently on short transcripts

The UL_{190} and US_{146} genes each encode transcripts ending at the U1 3′-end which are in the same size range as snRNAs. There are two discrete fragments protected by the probe which are derived from the transfected genes, the shorter one ending at the U1 3′-end and the longer one ending at the histone 3′-end. At least 98% of the transcripts from the UL_{190} gene transfected into CHO cells ended at the U1 3′-end (Fig. 2A, lane 2). Transcripts ending at the histone 3′-end were barely detectable only in long autoradiographic exposures.

In CHO cells transfected with the US_{146} gene, 80% of the transcripts in steady-state RNA had U1 3′-ends while only 20% of the transcripts had histone 3′-ends (Fig. 2B, lanes 1 and 3). Treatment of cells with actinomycin D for 45 min resulted in complete disappearance of the transcripts with U1 3′-ends, while the amount of transcripts with histone 3′-ends was only reduced 50% (Fig. 2B, lanes 2 and 4). Thus the transcripts ending at the U1 3′ box were much less stable than the transcripts ending at the histone 3′-end, probably because they lack the secondary structure at the 3′-end of the U1 snRNA which is present in the

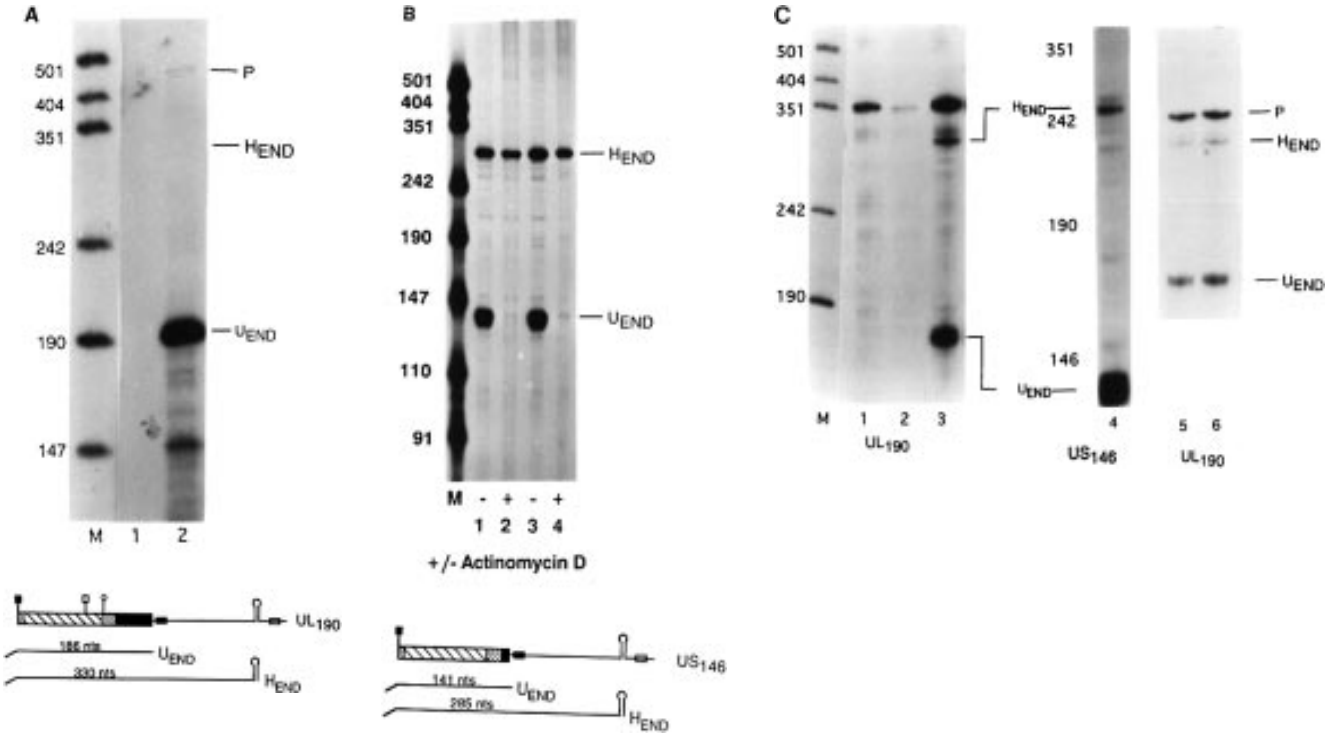

Figure 2. Expression of the UL_{190} genes and US_{146} genes. (**A**) 20 µg RNA from CHO cells transfected with the UL_{190} gene (lane 2) or 10 µg yeast tRNA (lane 1) were analyzed by S1 nuclease mapping using the UL_{190} gene labeled at the 3′-end of the *Xba*I site adjacent to the U1 coding region as probe. A diagram of the S1 nuclease assay is shown below the figure. The protected fragments are: U_{END}, protection to the U1 3′-end; H_{END}, protection of the 3′-end of the mouse histone H2a gene. P is the residual undigested probe. The position of the histone end is indicated and was barely detectable on the film. (**B**) 10 µg RNA from CHO cells transfected with the US_{146} gene were analyzed by S1 nuclease mapping using the US_{146} gene labeled at the 3′-end of the *Xba*I site. RNA was prepared from untreated duplicate cultures (lanes 1 and 3, –) and from cultures which had been treated for 45 min with 5µg/ml actinomycin D (lanes 2 and 4, +). Lane M is pUC18 digested with *Hpa*II. The protected fragments are as in (A). (**C**) The UL_{190} gene (lanes 3, 5 and 6) or the US_{146} gene (lane 4) were injected into *Xenopus* oocytes and the 3′-end of the transcripts analyzed by S1 nuclease mapping using the appropriate gene labeled at the 3′-end of the *Xba*I site as probe. In lane 5 the oocytes were injected with 0.15 ng DNA and in lane 6 the oocytes were injected with 0.45 ng UL_{190} DNA. The protected fragments are labeled as in (A). Lane 1 shows analysis of 10µg yeast tRNA and lane 2 is analysis of RNA from uninjected oocytes. The band at 350 nt in lanes 1–3 is derived from the probe.

UL genes. Transcripts from the US genes with U1 3′-termini may be under-represented in steady-state RNA.

Similar results were obtained when the UL_{190} and US_{146} genes were injected into *Xenopus* oocytes. Ninety percent of the transcripts from the UL_{190} gene end at the U1 3′-end and only 10% of the transcripts ended at the histone 3′-end (Fig. 2C, lane 3). When the US_{146} gene was injected into frog oocytes, again >90% of the transcripts ended at the U1 3′-end and <10% of the transcripts ended at the histone 3′-end (Fig. 2C, lane 4). This result was not changed by altering the amount of the UL_{190} gene injected by a factor of 3 (Fig. 2C, lanes 5 and 6), demonstrating that the relative proportion of different 3′-ends was not a result of overloading either the histone or U1 3′-end formation machinery.

Since the US_{146} gene has only 5 nt of U1 coding sequence at the 5′-end and 10 nt of U1 coding sequence at the 3′-end and lacks all secondary structure features of U1 RNA, these results demonstrate that efficient formation of U1 3′-ends does not require any sequences in the coding region, as suggested previously by Hernandez and Weiner (1). However, the last stem–loop structure may be important for stability of the transcripts.

snRNA 3′-ends are formed with moderate efficiency on transcripts of 280–400 nt

A series of genes yielding snRNA transcripts containing histone coding region and 280–400 nt in length were constructed, ending with either the last 49 nt of U1 snRNA (Fig. 1A) or the last 10 nt of U1 RNA (Fig. 1B). When we assayed the transcripts from these four genes in either CHO cells or in *Xenopus* oocytes, we observed that only 50–70% of the transcripts ended at the U1 3′-end (Fig. 3A).

About 60% of the transcripts from the UL_{355} gene ended at the U1 3′-end and 40% ended at the histone 3′-end (Fig. 3A). The stability of each of the transcripts was estimated by comparing the relative amounts of each transcript in exponentially growing cells and in cells treated for 1 h with actinomycin D. There was a similar reduction in the transcripts ending at the U1 3′- and the histone 3′-ends, demonstrating that these two transcripts had similar stabilities (Fig. 3A, lanes 2 and 3). A similar proportion of U1 and histone 3′-ends was seen when the US_{310} gene was transfected into CHO cells (Fig. 3A, lanes 4 and 5).

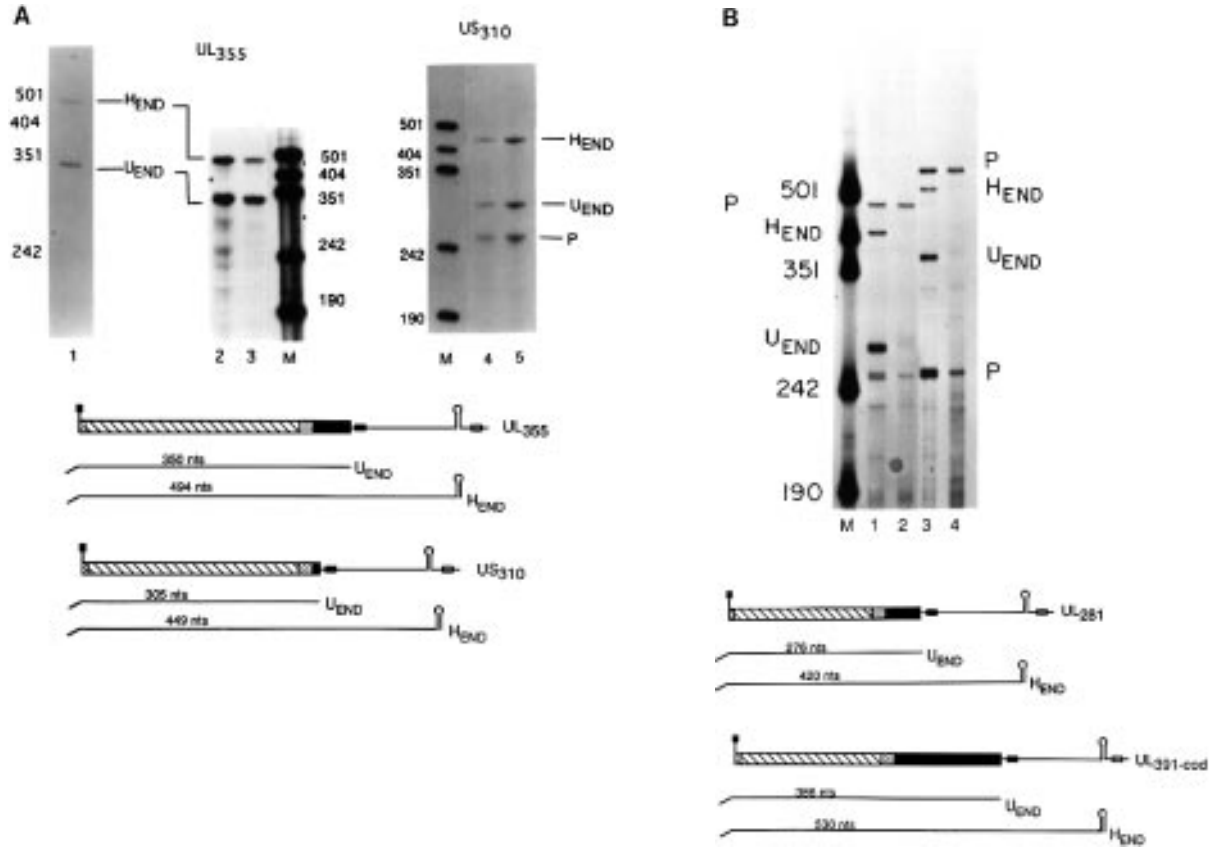

Figure 3. Expression of genes with intermediate transcript lengths. (**A**) 10 µg RNA from cells transfected with the UL_{355} gene (lane 1), using the gene labeled at the *Xba*I site as probe. Parallel cultures of cells, either control cells (lane 2) or cells treated with 5µg/ml actinomycin D for 1 h (lane 3), were analyzed. 5 or 10 µg RNA from cells transfected with the US_{310} gene (lanes 4 and 5) were analyzed by S1 nuclease mapping using the gene labeled at the *Xba*I site as probe. Lane M is pUC18 digested with *Hpa*II. The protected fragments are labeled as in Figure 2A. The fragment labeled P (lanes 4 and 5) is derived from the probe. (**B**) 10 µg RNA from CHO cells transfected with the UL_{281} gene (lane 1) or the $UL_{391\text{-}cod}$ gene (lane 3) were analyzed by S1 nuclease mapping using the appropriate gene labeled at the 3′-end of the *Xba*I site (in the polylinker 5 nt after the U1 sequence at the start of the gene) as probe. Lanes 2 and 4 are 10 µg RNA from untransfected CHO cells analyzed with the UL_{281} or $UL_{391\text{-}cod}$ probes respectively. The fragments in lanes 2 and 4 are derived from the probe. Lane M is pUC18 digested with *Hpa*II. The protected fragments are as in (A).

A similar distribution of transcripts with histone and U1 3′-ends was obtained with the UL_{281} gene, which lacks the entire 5′-end of the histone coding region, and the $UL_{391\text{-}cod}$ gene, which lacked the entire 3′-portion of the histone coding region (Fig. 3B, lanes 1 and 3). Note that the $UL_{391\text{-}cod}$ gene has the complete U1 sequence at the 3′-end of the transcript and the presence of the complete U1 RNA sequence was not sufficient to cause efficient termination at the U1 3′-end.

Similar results were obtained when these constructs were injected into *Xenopus* oocytes. About 60% of the transcripts from the UL_{355}, UL_{281} and $UL_{391\text{-}cod}$ genes ended at the U1 3′-end and 40% at the histone 3′-end (not shown). Thus, increasing the length of the transcribed region resulted in a decrease in the efficiency of formation of snRNA 3′-ends.

snRNA 3′-ends are formed inefficiently on long transcripts

Previously we showed that the UHU_L and UHU_S genes, which have a U1 promoter and 3′-end and a complete histone coding region, but do not have a histone 3′-end downstream, formed a small number of transcripts which ended at the U1 3′-end (4). Since these genes did not have a histone 3′ processing site downstream, it was not possible to determine the efficiency of U1 3′-end formation. To determine the efficiency of U1 3′-end formation on these long transcripts and to determine whether the presence of the distal histone 3′ processing signal affected usage of the U1 snRNA 3′-end, we compared expression of the UHU genes with the UL_{580} and US_{536} genes.

Two types of transcripts were also detected from the UHU genes: transcripts that start at the U1 start site and end at the U1 3′-end (536 nt from the UHU_S gene and 580 nt from the UHU_L gene) and read-through transcripts that extend into the 3′ flanking plasmid sequences. The read-through transcripts were detected using the UL_{580} or US_{536} gene as probe, mapping all transcripts which extended past the U1 3′-end as a single protected fragment (RT in Fig. 4A and B). The read-through transcripts have heterogeneous 3′-termini, since there were no discrete protected fragments extending past the U1 3′-end when the UHU genes were used as probes (4).

Figure 4A shows analysis of the transcripts formed from the UHU_L and UL_{580} genes in CHO cells. The UL_{580} gene was labeled at the 3′-end of the *Nar*I site (nt 204) in the histone coding region. The great majority of the transcripts from the UL_{580} gene ended at the histone 3′-end, with only 10% of the transcripts ending at the U1 3′-end (Fig. 4A, lane 3). Equal amounts of read-through transcripts and transcripts ending at the U1 end were present in steady-state RNA produced from the UHU_L gene (Fig. 4A, lane 1),

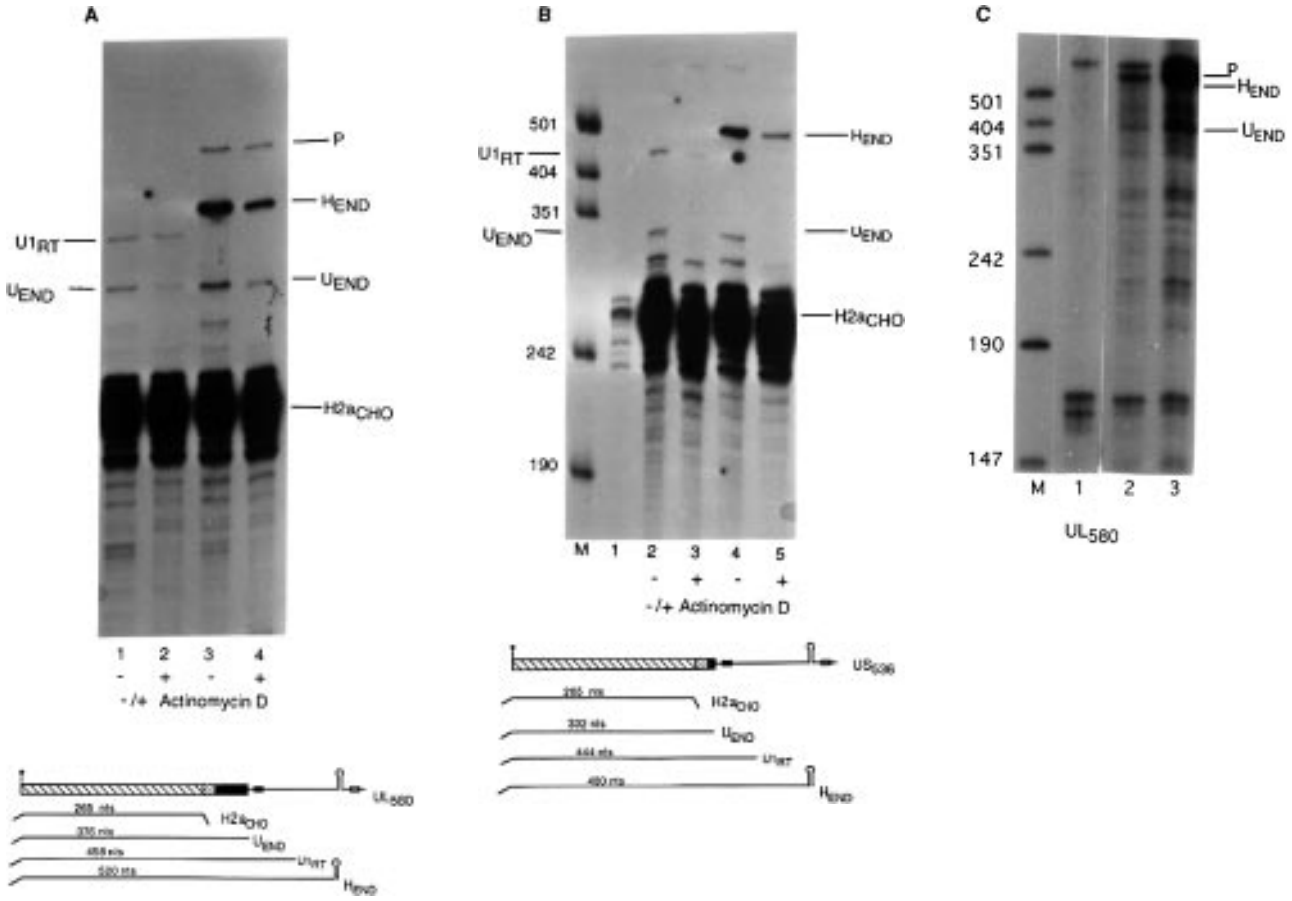

Figure 4. Expression of the UHU, UL_{580} and US_{536} genes. (**A**) 10 µg RNA from CHO cells transfected with the UHU_L gene (lanes 1 and 2) or the UL_{580} gene (lanes 3 and 4) were analyzed by S1 nuclease mapping using the UL_{580} gene labeled at the 3′-end of the *Nar*I site as probe. RNA was prepared from duplicate cultures, one a control culture (lanes 1 and 3, labeled –) and one which was treated with 5 µg/ml actinomycin D for 45 min (lanes 2 and 4, labeled +). Note that the UL_{580} probe will map all transcripts extending >82 nt past the U1 3′-end of UHU_L as a single protected fragment ($U1_{RT}$). The protected fragments are: $H2a_{CHO}$, protection of the endogenous hamster H2a genes; U_{END}, protection to the U1 3′-end; H_{END}, protection to the histone end; $U1_{RT}$, protection of transcripts which extend >82 nt past the U1 3′-end. The position of the undigested probe (P) is indicated. (**B**) 10 µg RNA from CHO cells transfected with the UHU_S gene (lanes 2 and 3) or the US_{536} gene (lanes 4 and 5) were analyzed by S1 nuclease mapping using the US_{536} gene labeled at the 3′-end of the *Nar*I site at codon 43 of the H2a sequence as probe. RNA was prepared from duplicate cultures, one a control culture (lanes 2 and 4, labeled –) and one which was treated with 5µg/ml actinomycin D for 45 min (lanes 3 and 5, labeled +). 10 µg RNA from untransfected CHO cells was analyzed in lane 1. Note that the US_{536} probe will map all transcripts extending >112 nt past the U1 3′-end of the UHU_S gene as a single protected fragment. The protected fragments are labeled as in (A). Lane M is pUC18 digested with *Hpa*II. Shorter exposures of the gels in (A) and (B) confirmed that the concentration of the endogenous histone mRNA was similar in all lanes. (**C**) Total RNA was extracted from *Xenopus* oocytes injected with the UL_{580} gene and analyzed by S1 nuclease mapping using the gene labeled at the 3′-end of the *Nar*I site at codon 45 of the histone coding region as probe (see A). Lane 1 is 3 µg RNA from uninjected oocytes. Lanes 2 and 3 are 2 and 6 µg RNA from the injected oocytes. The protected fragments are labeled as in (A). P is the residual undigested probe. Lane M is pUC18 digested with *Hpa*II.

in contrast to the 10% of the transcripts which end at the U1 3′-end from the UL_{580} gene.

When CHO cells transfected with the UL_{580} gene were treated with actinomycin D for 1 h, there was a loss of ~75% of the transcripts with the U1 3′-end and the histone 3′-end, consistent with a half-life of ~30 min for each transcript (Fig. 4A, cf. lanes 1 and 3 with lanes 2 and 4). Thus the transcripts ending at the U1 and histone 3′-ends have similar stabilities. In contrast, the small amount of read-through transcripts from the UHU_L gene present in steady-state RNA is more stable than the transcripts with the U1 ends. The read-through transcripts are derived from a population of transcripts which have heterogeneous 3′-ends and hence different half-lives and the read-through transcripts present in steady-state RNA represent the most stable of these transcripts.

Similar results were obtained when the transcripts from the UHU_S gene and US_{536} genes were analyzed. Approximately 50% of the transcripts from the UHU_S gene found in steady-state RNA end at the U1 3′-end and 50% are longer read-through transcripts. In contrast, >90% of the transcripts from the US_{536} gene end at the distal histone 3′-end. Again the absolute amount of transcripts with a U1 3′-end from the UHU_S and US_{536} genes is similar (Fig. 4B, lanes 2 and 4). Treatment of cells with actinomycin D showed that the transcripts which ended at the U_S 3′-end were less stable than those which ended at the U_L 3′-end (Fig. 4B, lanes 3 and 5; cf. with Fig. 4A). In longer exposures, transcripts ending at the U1 end in the actinomycin D-treated cells were detectable. Thus, in both the UL_{580} and US_{536} genes the great majority of transcripts end at the distal histone 3′-end instead of terminating at the proximal U1 end.

There are two possible explanations for the difference in the relative amounts of the U1 3′-ends from the UHU_L and UL_{580} genes. First, there could be a direct effect of the distal histone

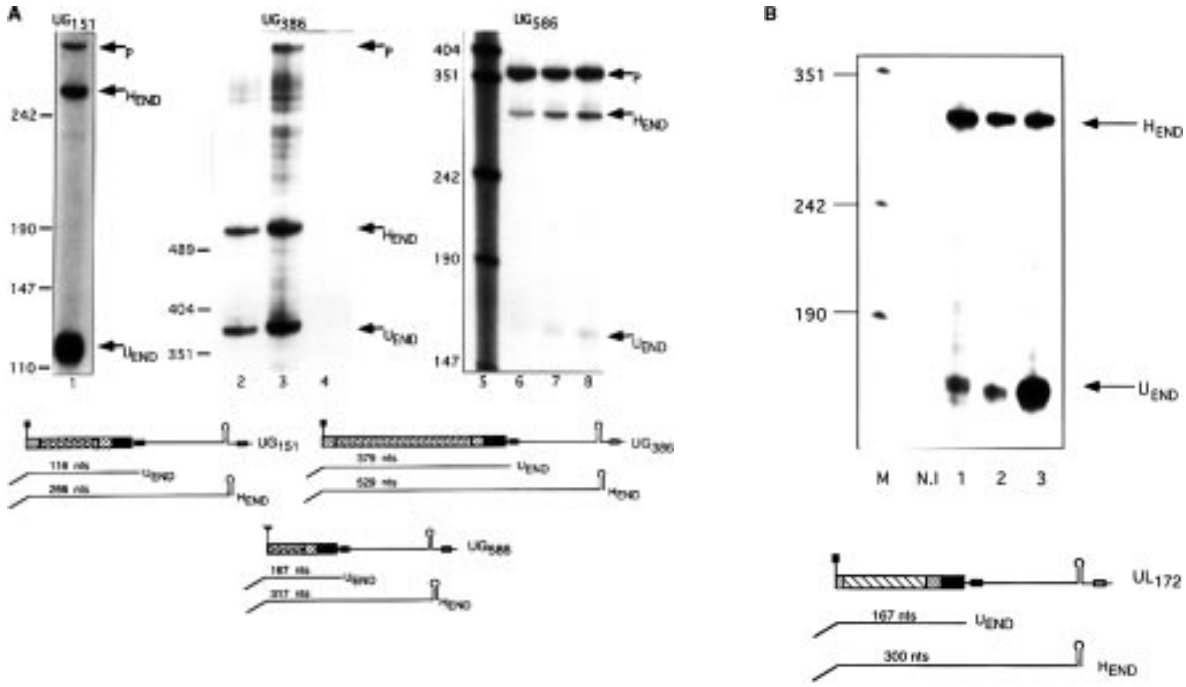

Figure 5. Expression of UG genes. (**A**) The UG_{151} gene (lane 1), the UG_{386} gene (lanes 2 and 3) or the UG_{586} gene (lanes 6–8) was injected into *Xenopus* oocytes and RNA prepared 18 h later. The RNAs were analyzed by S1 nuclease mapping using the gene labeled at the *Xba*I site at nt 5 in the transcript (UG_{151} gene, lane 1; UG_{386} gene, lanes 2–4) or the *Hin*dIII site in the globin coding region (UG_{586} gene, lanes 6–8). Three times as much RNA was analyzed in lane 3 compared with lane 2 and increasing amounts of RNA (1×, 2×, 4×) were analyzed in lanes 6–8. Lane 4 is analysis of 10 µg yeast tRNA. Lane 5 is pUC18 digested with *Hpa*II. The protected fragments are: U_{END}, protection to the U1 3′-end; H_{END}, protection to the histone end. The bands labeled P are the undigested probes. The S1 nuclease assays are illustrated below the figure. (**B**) The UG_{670} gene (lane 1), the UG_{350} gene (lane 2) or the UL_{172} gene (lane 3) was injected into *Xenopus* oocytes and RNA prepared 18 h later. These three genes share the same sequences at the 3′-end of the transcripts, allowing mapping of the transcripts from these genes using the UL_{172} gene labeled at the 3′-end of the *Xba*I site as probe. The protected fragments are labeled as in (A). Lane M is pUC18 digested with *Hpa*II and the lane N.I is analysis of RNA from uninjected oocytes.

3′-end on the efficiency of utilization of the U1 3′-end. Second, the great majority of the heterogeneous read-through transcripts from the UHU genes could be very unstable and not represented in steady-state RNA. We favor the second interpretation for the following reasons. Stable transfections of genes into CHO cells have yielded reproducible levels of expression of the transfected histone and U1 genes (16; unpublished results). When we compare the absolute amount of expression of the UL_{580} and UHU_L genes (by comparing the intensity of the protected fragments with the endogenous hamster histone mRNA), we observe that the absolute amount of U1 3′-ends formed from the UHU_L gene is not greater than the amount of transcript ending at the U1 end formed from the UL_{580} gene (Fig. 4A, cf. lanes 1 and 3). Thus the distal histone 3′-end does not reduce utilization of the U1 3′-end. It is very likely that most of the heterogeneous read-through transcripts are very unstable, accounting for the low amount of read-through transcripts from the UHU_L gene.

We also analyzed the transcripts formed from the UL_{580} and US_{536} genes after injection of the genes into *Xenopus* oocytes and obtained similar results. The UL_{580} gene, which has 580 nt between the start site and the U1 3′-end, gave primarily transcripts with histone 3′-ends and only a small number of transcripts which ended at the U1 3′-end (Fig. 4C, lanes 2 and 3). Similar results were found with the US_{536} gene (data not shown). Thus there is inefficient usage of the U1 3′ box on long transcripts in both mammalian cells and *Xenopus* oocytes.

Efficiency of snRNA 3′ end formation is sequence independent

To rule out an effect of histone coding sequences, a series of genes encoding short (UG_{151}), intermediate (UG_{350} and UG_{386}) and long (UG_{586} and UG_{670}) transcripts ending at the U1 3′-end were constructed (Fig. 1C). The UG_{151}, UG_{350} and UG_{586} genes have no sequences in the transcribed region in common with the UL and US genes, other than the 3′-end and the first 5 nt of the U1 coding region. These genes were injected into *Xenopus* oocytes and the proportions of the transcripts ending at the U1 and histone 3′-ends were measured. The results were similar to those obtained for the UL genes. Over 80% of the transcripts from the UG_{151} gene ended at the U1 3′-end (Fig. 5A, lane 1), ~50% of the transcripts from the UG_{386} gene ended at the U1 3′-end and 50% at the histone 3′-end (Fig. 5A, lanes 2 and 3) and only 15% of the transcripts from the UG_{586} gene ended at the U1 3′-end.

Similar results were obtained with the UG_{350} and UG_{670} genes. Since these genes are identical for the last 167 nt with the UL_{172} gene, the transcripts from these genes can all be mapped with a single probe, allowing direct comparison of the relative amounts of the transcripts ending at the U1 and histone 3′-ends. The great majority of the transcripts (90%) from the UG_{670} gene end at the distal histone 3′-end (Fig. 5B, lane 1), while only 10% of the transcripts end at the proximal U1 3′-end. In contrast, 90% of the transcripts from the UL_{172} gene ended at the U1 3′-end (Fig. 5B,

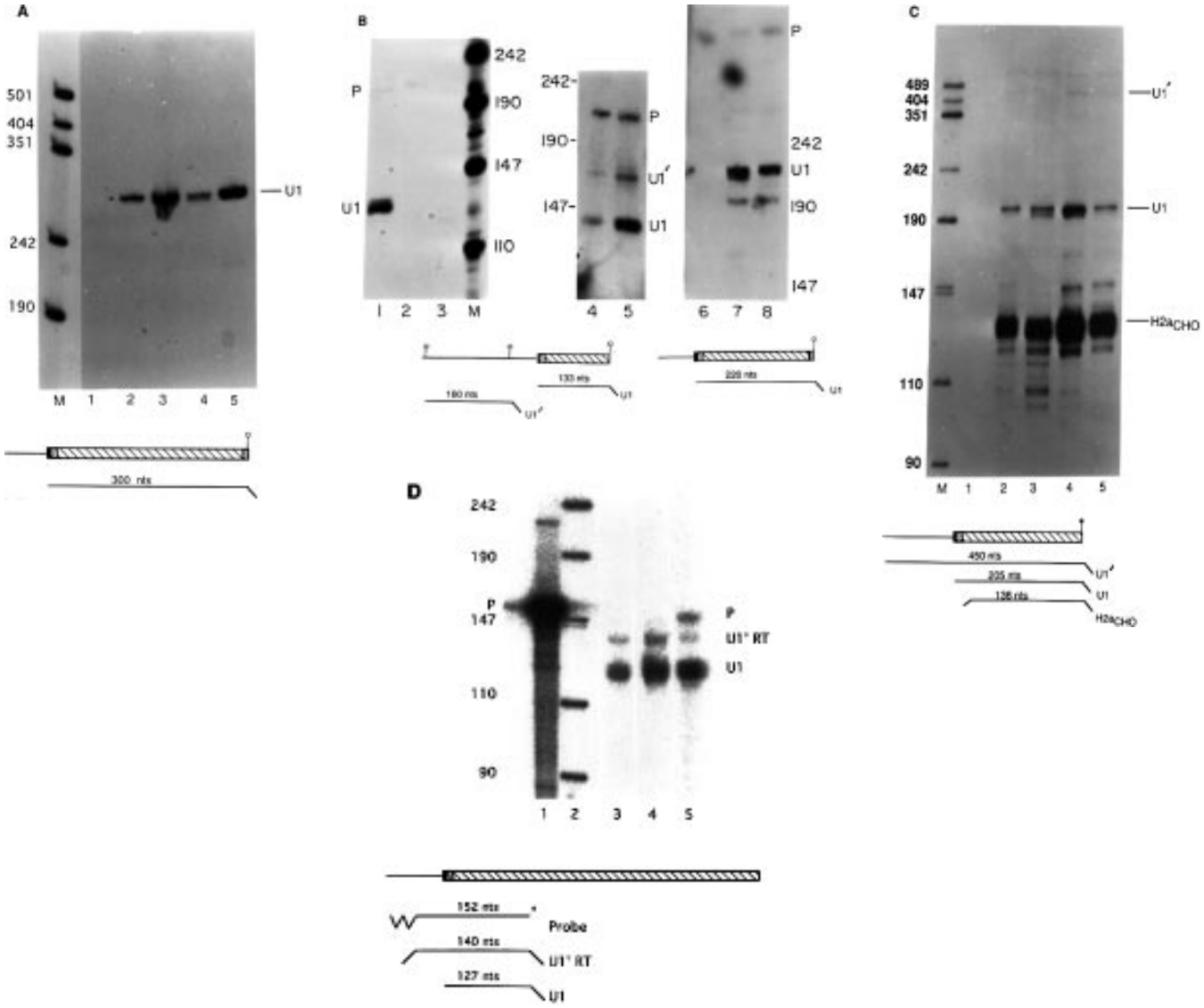

Figure 6. 5′-Ends of transcripts from the UHU, UL and US genes. (**A**) Total cell RNA from CHO cells transfected with the UL_{355} genes (lanes 2 and 3) or from CHO cells transfected with the US_{310} gene (lanes 4 and 5) was analyzed using a probe labeled at the 5′-end of the *Xma*I site adjacent to the U1 3′ sequences. 5 (lanes 2 and 4) or 20 µg (lanes 3 and 5) RNA were analyzed. Lane 1 is analysis of 10 µg yeast tRNA. The protected fragment (U1) maps to the start of the U1 coding sequence. (**B**) 5′-Ends of transcripts from the US_{146}, $UL_{391–cod}$ and UL_{281} genes. The 5′-ends of transcripts from the US_{146} and $UL_{391\text{-}cod}$ genes were mapped using a probe labeled at the 5′-end of the *Ava*I site at codon 20 of the H2a-614 gene. 10 µg total cell RNA from CHO cells transfected with the US_{146} gene (lane 1), untransfected CHO cells (lane 2) or yeast tRNA (lane 3) were analyzed. 2 (lane 4) or 10 µg (lane 5) total cell RNA from CHO cells transfected with the $UL_{391\text{-}cod}$ gene were analyzed by S1 nuclease mapping using the US_{146} gene labeled at the 5′-end of the *Ava*I site as probe. The band labeled U1 represents protection to the first nucleotide of U1 RNA. The band labeled U1′ indicates transcripts which initiated upstream of the U1 promoter. This arises from protection of a probe fragment labeled at the *Ava*I site at 90 nt before the U1 start site and the protected fragment maps to the same site observed for the UHU, UL and US genes in Figure 2C. P indicates the position of the undigested probe. Lane M is marker pUC18 digested with *Hpa*II. In lanes 6–8 a probe labeled at the 5′-end of the *Xma*I site in the polylinker of the UL_{281} gene just before the U1 3′-end was used. 10 µg yeast tRNA (lane 6) or 10 µg total cell RNA from duplicate cultures of CHO cells transfected with the UL_{281} gene (lanes 7 and 8) were analyzed. The 228 nt fragment (labeled U1) is due to protection to the first nucleotide of the U1 RNA. The 450 nt fragment labeled P is the undigested probe. The origin of the band at 195 nt is not known, although it was not observed in all analyses and is probably an S1 nuclease artifact. A diagram of the S1 nuclease assays is shown below the figure. (**C**) 5′-Ends of transcripts from the UHU, UL_{580} and US_{536} genes. 10 µg total RNA from cells transfected with the US_{536} (lane 2), UL_{580} (lane 3), UHU_L (lane 4) or UHU_S (lane 5) genes were analyzed using an S1 nuclease protection assay. The 5′-ends were mapped using the U_5H gene labeled at the 5′-end of the *Nar*I site as probe. This probe also maps longer transcripts, which initiated 5′ of these genes as a 450 nt fragment (4). These transcripts, labeled U1′, represent <5% of the total transcripts. Lane 1 is analysis of 10 µg yeast tRNA. The other protected fragments are: U1, protection to the start of the U1 coding sequence; $H2a_{CHO}$, protection to the ATG codon by endogenous hamster H2a mRNA. Lane M is marker pUC18 digested with *Hpa*II. (**D**) 5′-Ends of transcripts from the UL_{580} and $UL_{391–cod}$ genes. A 152 nt probe which contained 12 nt of heterologous DNA, 13 nt of the U1 promoter and 127 nt of coding sequence. The probe was labeled at the 5′-end and used for S1 nuclease mapping of total RNA from CHO cells transfected with the UL_{580} or $UL_{391–cod}$ genes. A diagram of the S1 nuclease assay is given below the figure. The 127 nt fragment (U1) results from protection of transcripts initiated at the U1 start site and the 140 nt fragment ($U1_{RT}$) results from protection of any transcripts which initiated >13 nt upstream of the U1 promoter. Lane 1, the 152 nt probe; lane 2, pUC18 digested with *Hpa*II; lanes 3 and 4, 25 and 50 µg RNA from cells transfected with the $UL_{391–cod}$ gene; lane 5, 50 µg RNA from the UL_{580} gene.

lane 3). The UG_{350} gene gave an intermediate result, with only 30–40% of the transcripts ending at the U1 3′-end (Fig. 5B, lane 2). We conclude that the length dependence of snRNA 3′-end formation is independent of the sequences in the transcript.

Almost all the transcripts initiate at the U1 5′-end

One possible explanation for these results could be that the transcripts which end at the histone 3′-end are not initiated from the U1 snRNA promoter, but from some cryptic promoter. Transcripts initiating at a cryptic promoter upstream of the U1 start site do not direct U1 3′-end formation (4). To rule out the possibility that formation of histone 3′-ends was a result of initiation from a cryptic promoter, we have mapped the 5′-ends of the transcripts from all of the genes. Figure 6A shows the 5′-ends of the transcripts from the UL_{355} and US_{310} genes, mapped using a probe labeled just prior to the spot where the U1 3′-end was attached. A single protected fragment of 300 nt, the expected length for transcripts initiating at the U1 5′-end, was observed with both of these genes (Fig. 6A, lanes 2–5). Figure 6B shows the 5′-ends of the transcripts from the other genes. The transcripts were mapped using probes labeled at the 5′-end of the *Ava*I site at codon 20 of the H2a-614 coding region in the $UL_{391\text{-}cod}$ gene. The great majority of transcripts from these genes initiated at the U1 snRNA start site (Fig. 6B, lanes 1, 4, 5, 7 and 8). A small amount of transcripts (labeled U1′) was detected which initiated upstream of the U1 promoter (Fig. 6B, lanes 4 and 5; fragment U1′). No upstream starts were detected from the US_{146} or UL_{281} genes (Fig. 6B, lanes 1, 7 and 8).

Figure 6C shows the analysis of the 5′-ends of the transcripts from the UHU_L, UHU_S, US_{536} and UL_{580} genes, using an S1 nuclease assay with a probe which is labeled at the 5′-end of the *Nar*I site (codon 45) of the histone H2a gene. More than 95% of the transcripts map to the U1 start site. There is only a small amount of transcripts which initiate ~200 nt 5′ of the U1 start site (labeled U1′ in Fig. 6C).

The assays in Figure 6A–C rule out the presence of large amounts of transcripts initiating at a defined site 5′ of the gene. However, they do not rule out the possibility of a heterogeneous set of transcripts initiating upstream of the U1 promoter. These would not have been detected in the previous assay, since they would not map a defined protected fragment. To assess the amount of transcripts which initiated upstream of the U1 promoter, we constructed a probe which contained 12 nt of heterologous sequence, 13 nt of the U1 promoter and 127 nt of coding region. This probe will map all the transcripts which come from upstream of the U1 start site as a single fragment 12 nt longer than the properly initiated fragment. The ratio of the two protected fragments gives the relative amount of properly initiated transcripts. The fragments were quantified on a PhosphorImager. Ninety five percent of the transcripts from the UL_{580} gene initiated at the U1 start site (Fig. 6D, lane 5). Since the great majority of the transcripts from the UL_{580} gene ended at the histone 3′-end, most of the transcripts which ended at the histone 3′-end must have initiated at the U1 start site. Similarly, >85% of the transcripts from the $UL_{391\text{–}cod}$ gene were initiated correctly (Fig. 6D, lanes 3 and 4), in agreement with the results in Figure 6B (lanes 4 and 5). Taken together these results indicate that there were not significant amounts of improperly initiated transcripts which contributed to these results.

DISCUSSION

The mechanism of formation of 3′-ends of the U series of snRNAs in vertebrates is unique among genes transcribed by RNA polymerase II. First, there is only a single sequence element which lies 3′ of the snRNA sequence required for 3′-end formation (3,22). This is in contrast to the bipartite elements required for 3′-end formation found in both histone and polyadenylated mRNAs, which define a cleavage site located between them (23). Second, snRNA 3′-end formation in vertebrates is tightly coupled to transcription initiation and there is an absolute requirement for initiation from an snRNA promoter (1,2,24,25). The initial transcript from the snRNA genes is formed by transcription termination.

We have previously shown that the histone H2a sequence does not contain any cryptic U1 3′ box signals (4) and that the U1 promoter can efficiently drive expression of the histone mRNA and that the histone 3′-end is formed efficiently on these transcripts (4). Previously we showed that the U1 promoter expresses the α-globin protein efficiently in CHO cells (26). In the results reported here, we define a length requirement for efficient coupling of the U1 promoter to U1 3′-end formation. Since the human α-globin coding sequence and the histone H2a sequence are totally dissimilar, this phenomenon is independent of the sequences transcribed and dependent solely on the length of the transcribed region. These results depend on the transcripts initiating at the U1 promoter, since transcripts which read through these genes which initiated elsewhere would all end at the histone 3′-end. The great majority of all the transcripts initiated at the proper U1 start site. Similar results have been seen with genes expressing β-galactosidase from these promoters (26).

The relative usage of the U1 and histone 3′-ends as a function of the length of the U1 transcripts in both CHO cells and *Xenopus* oocytes is summarized in Figure 7. U1 3′-ends are formed very efficiently on transcripts <200 nt, the size range of the vertebrate snRNAs. U1 3′-ends are formed with reduced efficiency (50–70%) on transcripts between 280 and 400 nt and are formed very inefficiently on transcripts >500 nt.

Mechanism of coupling the promoter to 3′-end formation

Three features of 3′-end formation of snRNAs in vertebrates are a strict dependence on initiation from an snRNA promoter (1,2,24,25), a dependence on the length of the transcribed region (this paper) and that the 3′-end signal is not an essential part of the promoter, since absence of the 3′ signal does not affect the level of expression from the U1 promoter (4). The mechanism coupling transcription initiation to 3′-end formation of the snRNA genes presumably evolved to allow efficient expression of these small transcripts. We note that the yeast *Saccharomyces cerevisiae* snRNAs, which are often much longer than vertebrate snRNAs, ranging in size from 1175 to 106 nt (27,28), are transcribed from polymerase II promoters similar to mRNA promoters and are transcribed efficiently from mRNA promoters (27,29). Thus it is likely that 3′-end formation of yeast snRNAs is not dependent on the length of the transcribed region. The absolute requirement for transcription from an snRNA promoter to form snRNA 3′-ends may be unique to vertebrates. There is not a strong coupling of the snRNA promoter to formation of the 3′-ends of invertebrate (sea urchin) snRNAs (30) or of plant snRNAs (31), although these RNAs are also the size of the vertebrate snRNAs. The 3′-end of sea urchin snRNAs, like the 3′-end of vertebrate snRNAs, is formed co-transcriptionally (30) and hence there is likely to be a similar length dependence for 3′-end formation of these RNAs.

There are two possible mechanisms for coupling transcription termination with the promoter. First, a factor could bind the 3′-end

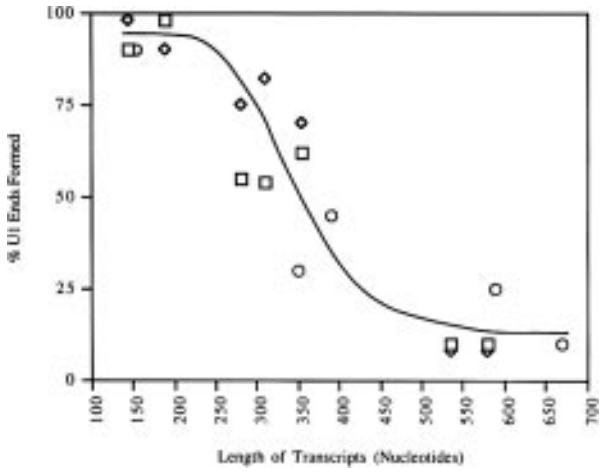

Figure 7. Length dependence of formation of UI 3 ends. The percentage of transcripts ending at the UI 3-end is plotted as a function of the lengths of the transcripts ending at the UI end. The data include analysis of the genes in both mammalian cells and Xenopus oocytes. The data were obtained by densitometry of the autoradiographs or from analysis on a PhosphorImager. The squares are the results for the UL and US genes in CHO cells, the diamonds the results for the UL and US genes in Xenopus oocytes and the circles the data for the UG genes in Xenopus oocytes.

signal in the DNA and then associate with the transcription initiation complex, presumably recognizing an essential component of the complex unique to snRNA genes (e.g. the factor which binds the proximal sequence element). Alternatively, a termination factor, which can specifically recognize the 3′ signal, could associate with the transcription complex during initiation, remain with the polymerase during elongation and stimulate termination when the 3′ signal is reached. The association between the factor and the transcription complex may be weak and the termination factor might dissociate (or be displaced) from the transcription complex on a long transcript before the 3′-end signal (either as RNA or DNA) is reached. Either of these mechanisms would be consistent with the requirement of the U1 snRNA promoter for 3′-end formation.

Recently Price and co-workers have shown that there is a transition during *in vitro* transcription which converts the transcription complex from one that pauses and/or terminates readily into a complex that is highly processive and resistant to many pause sites (32,33). It is possible that the length dependence of U1 snRNA 3′-end formation is a result of the transcription complex undergoing a transition to a stable, 'committed' state, refractory to termination. Prior to reaching this length, transcription can terminate readily at the U1 3′-end signal. While this explanation could account for the length dependence, it fails to account for the coupling of 3′-end formation to the U1 promoter. The snRNA promoter may promote the initial assembly of a transcription complex which is particularly prone to terminating at the snRNA 3′ box, while transcription complexes assembled on other promoters read through the 3′ box sequences readily.

We have observed inefficient U1 3′-end formation on U1 transcripts synthesized in isolated nuclei from mouse myeloma cells consistent with the possibility that the termination factor is easily lost during cell fractionation (34). Only if nuclei are prepared in such a way as to minimize loss of nuclear components (35) is there efficient coupling of transcription and 3′-end formation observed in transcripts synthesized in isolated nuclei. Taken together these results suggest that there is a *trans*-acting factor required for snRNA 3′-end formation associated with the transcription complex, which is readily lost from the complex. The biochemical basis of snRNA 3′-end formation remains to be elucidated.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM27789 to W.F.M and a grant from the Muscular Dystrophy Association to B.K.K. T.C.I. was supported by NIH Training Grant GM07092 to the Curriculum in Genetics and Molecular Biology.

REFERENCES

1 Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249–258.
2 Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
3 Ach,R.A. and Weiner,A.M. (1987) *Mol. Cell. Biol.*, **7**, 2070–2079.
4 Pilch,D.R. and Marzluff,W.F. (1991) *Gene Expression*, **1**, 41–53.
5 Neuman de Vegvar,H.E. and Dahlberg,J.E. (1990) *Mol. Cell. Biol.*, **10**, 3365–3375.
6 Murphy,J.T., Burgess,R.R., Dahlberg,J.E. and Lund,E. (1982) *Cell*, **29**, 265–274.
7 Kleinschmidt,A.M. and Pederson,T. (1987) *Mol. Cell. Biol.*, **7**, 3131–3137.
8 Yang,H., Moss,M.L., Lund,E. and Dahlberg,J.E. (1992) *Mol. Cell. Biol.*, **12**, 1553–1560.
9 Graves,R.A., Wellman,S.E., Chiu,I.-M. and Marzluff,W.F. (1985) *J. Mol. Biol.*, **183**, 179–194.
10 Hurt,M.M., Chodchoy,N. and Marzluff,W.F. (1989) *Nucleic Acids Res.*, **17**, 8876
11 Marzluff,W.F., Brown,D.T., Lobo,S. and Wang,S.S. (1983) *Nucleic Acids Res.*, **11**, 6255–6270.
12 Moussa,N.M., E-L-Din,A.S., Lobo,S.M. and Marzluff,W.F. (1987) *Nucleic Acids Res.*, **15**, 3622
13 Liu,T.-J., Levine,B.J., Skoultchi,A.I. and Marzluff,W.F. (1989) *Mol. Cell. Biol.*, **9**, 3499–3508.
14 Sun,J.-H., Pilch,D.R. and Marzluff,W.F. (1992) *Nucleic Acids Res.*, **20**, 6057–6066.
15 Chaney,W.G., Howard,D.R., Pollard,J.W., Sallustio,S. and Stanley,P. (1986) *Somatic Cell Mol. Genet.*, **12**, 237–244.
16 Levine,B.J., Liu,T.-J., Marzluff,W.F. and Skoultchi,A.I. (1988) *Mol. Cell. Biol.*, **8**, 1887–1895.
17 Kay,B.K. (1991) *Methods Cell Biol.*, **36**, 663–669.
18 Evans,J.P. and Kay,B.K. (1991) *Methods Cell Biol.*, **36**, 133–148.
19 Bentley,D.L. and Groudine,M. (1988) *Cell*, **53**, 245–256.
20 Spencer,C.A. and Groudine,M. (1990) *Oncogene*, **5**, 777–786.
21 Middleton,K.M. and Morgan,G.T. (1990) *Mol. Cell. Biol.*, **10**, 727–735.
22 Hernandez,N. (1985) *EMBO J.*, **4**, 1827–1837.
23 Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
24 Hernandez,N. and Lucito,R. (1988) *EMBO J.*, **7**, 3125–3134.
25 Parry,H.D., Tebb,G. and Mattaj,I.W. (1989) *Nucleic Acids Res.*, **17**, 3633–3644.
26 Bartlett,J.S., Sethna,M., Ramamurthy,L., Gowen,S.A., Samulski,R.J. and Marzluff,W.F. (1996) *Proc. Natl. Acad. Sci. USA*, **93,** 8852–8857.
27 Guthrie,C. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Ribonucleoprotein Particles*. Springer-Verlag, Berlin, Germany, pp. 196–212.
28 Roiha,H., Shuster,E.O., Brow,D.A. and Guthrie,C. (1989) *Gene*, **82**, 137–144.
29 Patterson,B. and Guthrie,C. (1987) *Cell*, **49**, 613–624.
30 Wendelburg,B.W. and Marzluff,W.F. (1992) *Mol. Cell. Biol.*, **12**, 4132–4141.
31 Connelly,S. and Filipowicz,W. (1993) *Mol. Cell. Biol.*, **13**, 6403–6415.
32 Kephart,D.D., Marshall,N.F. and Price,D.H. (1992) *Mol. Cell. Biol.*, **12**, 2067–2077.
33 Marshall,N.F. and Price,D.H. (1992) *Mol. Cell. Biol.*, **12**, 2078–2090.
34 Lobo,S.M. and Marzluff,W.F. (1987) *Mol. Cell. Biol.*, **7**, 4290–4296.
35 Lund,E. and Dahlberg,J.E. (1989) *EMBO J.*, **8**, 287–292.