The EMBO Journal Vol.18 No.10 pp.2867–2877, 1999

Transcription of the human U2 snRNA genes continues beyond the 3′ box *in vivo*

Paula Cuello, Diana C.Boyd, Michael J.Dye, Nick J.Proudfoot and Shona Murphy[1]

Chemical Pathology Unit, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

[1]Corresponding author
e-mail: smurphy@worf.molbiol.ox.ac.uk

The 3′ box of the human class II snRNA genes is required for proper 3′ processing of transcripts, but how it functions is unclear. Several lines of evidence suggest that termination of transcription occurs at the 3′ box and the terminated transcript is then a substrate for processing. However, using nuclear run-on analysis of endogenous genes, we demonstrate that transcription continues for at least 250 nucleotides beyond the 3′ box of the U2 genes. Although *in vivo* footprinting analysis of both the U1 and U2 genes detects no protein–DNA contacts directly over the 3′ box, a series of G residues immediately downstream from the 3′ box of the U1 gene are clearly protected from methylation by dimethylsulfate. In conjunction with the 3′ box of the U1 gene, this *in vivo* footprinted region causes termination of transcription of transiently transfected U2 constructs, whereas a 3′ box alone does not. Taken together, these results indicate that the 3′ box is not an efficient transcriptional terminator but may act as a processing element that is functional in the nascent RNA.

Keywords: pol II/RNA processing/snRNA/termination/transcription

Introduction

The vertebrate class II snRNA genes encode short non-polyadenylated nuclear RNAs, including the U1, U2, U4 and U5 snRNAs that are required for pre-mRNA processing. These genes are transcribed by RNA polymerase II (pol II) to give 3′-extended precursors that undergo processing themselves to produce the mature RNA components of snRNPs. Precursors with up to 16 nucleotide 3′ end extensions can be detected, and these undergo cap trimethylation, internal base modification and 3′ end trimming after association with U snRNP proteins in the cytoplasm (recently reviewed by Huang *et al*., 1997). The processed RNA and associated proteins are then transported back into the nucleus where assembly of the functional snRNP is completed (reviewed in Terns *et al*., 1993; Nagai and Mattaj, 1994). 3′ End formation of mature snRNAs requires a *cis*-acting sequence, known as the 3′ box, located 9–19 nucleotides downstream from the coding region of these snRNA genes (Hernandez, 1985; Yuo *et al*., 1985; Ciliberto *et al*., 1986; Neuman de Vegvar *et al*., 1986; Ach and Weiner, 1987; reviewed in Hernandez, 1992). Mutation of this sequence causes the accumulation of transcripts that are not processed accurately and are therefore longer than the mature snRNA.

Interestingly, the 3′ box functions efficiently only in the context of a pol II-dependent snRNA promoter. Although these genes are transcribed by the same polymerase as mRNA genes, their promoters are organized differently. Class II snRNA genes have characteristic short, compact, TATA-less promoters containing a distal sequence element (DSE) and a proximal sequence element (PSE) (reviewed by Hernandez, 1992; Lobo and Hernandez, 1994). Replacement of a U1 or U2 promoter with an mRNA promoter inhibits accurate 3′ end formation of the transcripts (Hernandez and Weiner, 1986; Neuman de Vegvar *et al*., 1986).

How the 3′ box functions is unknown, but several findings suggest that it may act as a transcription terminator: transcription of the human U1 genes terminates very close to the 3′ box (Kunkel and Pederson, 1985); only very small amounts of long snRNA gene transcripts containing a copy of the 3′ box have been detected (Hernandez, 1985; Kunkel and Pederson, 1985; Ach and Weiner, 1987; Lobo and Marzluff, 1987) and these do not appear to be precursors for processed RNA either in HeLa cells (Lobo and Marzluff, 1987) or in *Xenopus* oocytes (Ciliberto *et al*., 1986). Thus, an attractive hypothesis is that transcription terminates just upstream of the 3′ box and this produces a transcript that is a substrate for further processing, in analogy to the mechanism of 3′ end formation of vertebrate pre-rRNA (Kuhn and Grummt, 1989). However, a more direct role for this element in processing of snRNAs has not been ruled out and, in yeast, the 3′ ends of U1, U2 and U5 RNAs are produced by nuclease cleavage of longer precursors (Chanfreau *et al*., 1997; Elela and Ares, 1998; Seipelt *et al*., 1999).

In order to understand the exact role of the 3′ box, we have undertaken a study of transcriptional termination of the U2 snRNA genes using nuclear run-on analysis. We find that, in contrast to the U1 gene, transcription of the U2 gene continues for at least 250 bp after the 3′ box. Transcripts derived from the region downstream of the 3′ box are contiguous with RNA from upstream and are detected at very early points in a run-on time course, strongly suggesting that the profile we see reflects the *in vivo* polymerase loading.

A complementary *in vivo* footprinting analysis of the 3′ box region of the U1 and U2 genes also gave surprising results. No protection is detected over the 3′ box of either gene, whereas a clear area of protection from methylation by dimethylsulfate (DMS) is apparent immediately downstream from the 3′ box of the U1 gene.

These results suggest that termination of transcription may be occurring close to the 3′ box in the U1 genes due

to protein(s) binding to this DNA element. We provide direct evidence for this by showing that this region terminates transcription from a U2 promoter when placed downstream of a 3′ box. In contrast, a 3′ box alone does not function as an efficient terminator.

Taken together, these results indicate that most nascent snRNA gene transcripts contain a full copy of the 3′ box and transcription terminates beyond this element. We consider it likely, therefore, that the 3′ box is an RNA element that is required, in conjunction with a promoter-dependent activity, for rapid cleavage to produce the 3′ extended RNAs that are detectable *in vivo*.

Results

Transcription continues beyond the 3′ box of the U2 genes

A precise determination of the site of termination of transcription of the snRNA genes is crucial to an understanding of the role of the 3′ box in this process. We have therefore undertaken a detailed study of the polymerase loading on the human U2 genes using 47–50mer oligonucleotide probes in a nuclear run-on analysis.

There are ~10 U2 genes in the human haploid genome at 17q21–22 (Hammarstrom *et al*., 1984; Lindgren *et al*., 1985a; Pavelitz *et al*., 1995). These are tandemly repeated and the sequence of each 5.8–6.1 kb repeat unit is highly conserved (Van Arsdell and Weiner, 1984; Westin *et al*., 1984; Pavelitz *et al*., 1995), allowing us to analyse the distribution of polymerases over all of these genes using one set of probes. The position of each probe relative to the U2 gene and the results of a representative run-on analysis are shown in Figure 1A. A 47mer oligonucleotide complementary to probe A was included on the filters as a negative control (AS) and a 50mer oligonucleotide complementary to the pol III-transcribed 7SK RNA (7SK) was included as a positive control. The hybridization efficiency of the probes US to Z was determined using a labelled synthetic RNA (see Materials and methods), and the nuclear run-on results were normalized to these values after subtraction of α-amanitin-resistant transcription (see Figure 3).

No signal was detected over probe B in either the run-on or the synthetic RNA control, possibly due to the relatively high A/T content of this sequence, precluding analysis of this region. Probe B was included in subsequent run-on assays as an additional background control. The profile of polymerase density over the U2 genes determined by several such experiments is shown in Figure 1B as a percentage of the level of signal over probe C.

As expected, the polymerase density is low upstream of the site of initiation of the U2 genes (probe US, 2.65%), indicating that transcription from upstream genes does not continue into these promoters. However, the pattern of polymerase density over the U2 genes is strikingly different from that seen on the U1 genes where transcription was found to terminate close to the 3′ box (Kunkel and Pederson, 1985). Instead, a high level of polymerase loading appears to continue until at least 250 bp past the 3′ box (probe J, 95%). The polymerase density over region Z, ~600 bp downstream of the 3′ box, has dropped to 18%. We have not scored the polymerase density between the regions J and Z since this contains a $(CT)_n$ stretch which is not as highly conserved between individuals as the other sequences in the U2 genes (Liao and Weiner, 1995). These results suggest that a large proportion of transcription termination occurs downstream of the 3′ box in the U2 genes, perhaps mainly in the region between J and Z.

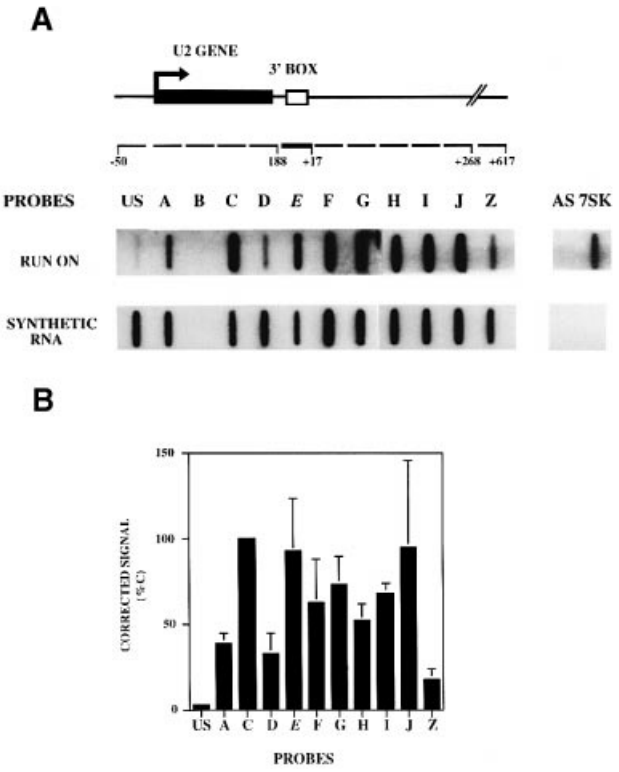

Fig. 1. Transcription continues beyond the 3′ box of the U2 genes. (**A**) A diagram of the structure of the U2 gene is shown with the relative positions of the probes marked below. The numbers noted next to the probes indicate the start or end of the probes used relative to the site of initiation (US and D) or relative to the end of the 3′ box (E, J and Z). The results of a run-on analysis and hybridization of synthetically produced RNA to the probes are shown below each probe. (**B**) A graphic representation is shown of the results of the run-on analysis in (A) as a percentage of the signal over probe C after subtraction of α-amanitin-resistant transcription and correction for the hybridization efficiency of each probe.

Transcripts downstream of the 3′ box are contiguous with the transcripts from the coding region and are detected early in a run-on time course

Although the pattern of run-on transcripts over the U2 gene suggests that a significant number of polymerases continue transcription beyond the 3′ box, we cannot exclude the possibility that transcription has reinitiated downstream of the coding region. We have therefore used an RNA probe complementary to region A–D of the U2 gene to hybrid-select transcripts contiguous with the coding RNA. A biotin-labelled selection probe was hybridized to the run-on RNA, and avidin-coupled magnetic beads were used to separate the probe and associated RNA from the mix (see Materials and methods). The selected RNA was then hydrolysed and hybridized to filters containing probes US to Z, AS and 7SK (Figure 2A). Importantly, transcripts from the 7SK gene have not been selected, indicating that the selection is specific.

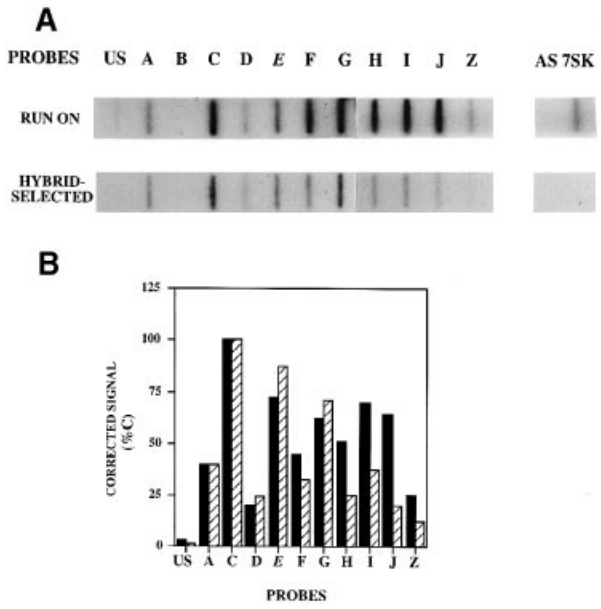

Fig. 2. Transcripts downstream of the U2 3′ box are contiguous with the transcripts from the coding region. (**A**) The results of hybridization of RNA that has been hybrid-selected after the run-on are shown below the profile from an unselected run-on (a lower exposure of the same experiment shown in Figure 1A). (**B**) A graphic representation of the results from an unselected run-on analysis (black bars) is shown compared with the results with RNA hybrid-selected after the run-on (cross-hatched bars), corrected as in Figure 1.

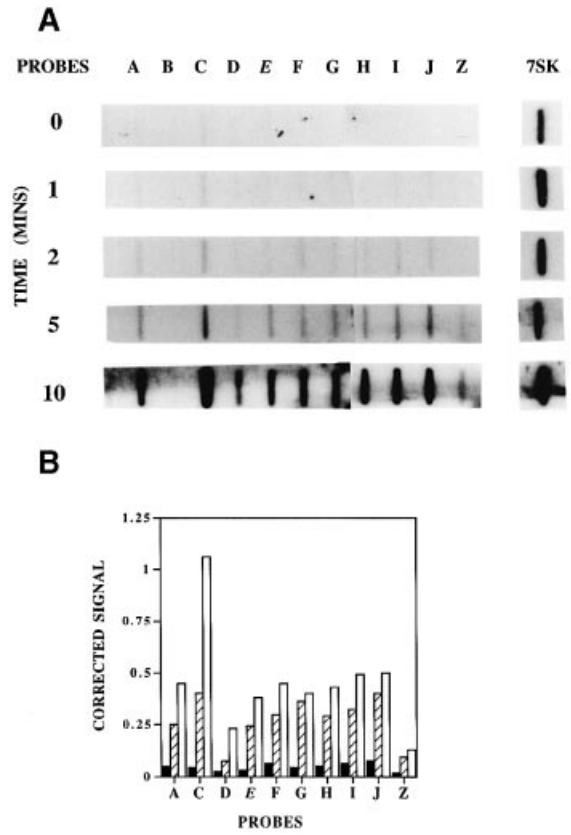

Fig. 3. Transcripts downstream of the U2 3′ box are detected early in a run-on time course. (**A**) The results of adding α-amanitin (2 μg/ml) to the run-on reaction at different times are shown. The time (in minutes) is shown at the left of the figure. The result of hybridization to the 7SK probe is shown for each time point at the right. (**B**) A graphic representation of the results corrected for the 7SK signal is shown for the 2 min (black bars), 5 min (cross-hatched bars) and 10 min (empty bars) time points.

Figure 2B shows the pattern of hybridization from this analysis (cross-hatched bars) compared with that seen with an unselected run-on analysis performed at the same time (black bars).

After hybrid selection, the signal over US relative to C has fallen from 2.6% to 1%, while the relative signals over probes A–G have remained at approximately the same levels. At least 50% of the transcripts complementary to probes H–I are also selected and are therefore contiguous with transcripts from the coding region. Some of the transcripts detected over probes J and Z are also contiguous with the coding RNA. The drop in signals over probes H–Z after hybrid selection indicates that some of the apparent transcription from this region may be the result of cross-hybridization or reinitiation. Alternatively, post-transcriptional nuclease digestion (specific or non-specific) may be responsible. For example, some of the transcripts detected over probes H–Z in the unselected run-on analysis may have been cleaved specifically in a 3′ box-dependent reaction.

The above results indicate that transcription is not terminating efficiently at the 3′ box in our run-on assay. However, transcription potentially can extend >1000 bp in the 15 min of a run-on reaction (Weber *et al.*, 1977). In addition, polymerases that are normally paused may resume transcription under run-on conditions (see, for example, Strobl and Eick, 1992). We have therefore carried out a time course of the run-on reaction to investigate the possibility that polymerases are not transcribing the region downstream of the 3′ box in intact cells (Figure 3).

α-Amanitin was added to 2 μg/ml, either before or 1, 2, 5 or 10 min after the addition of nucleotides, to the resuspended nuclei (see Materials and methods) (Figure 3A). The probe for the pol III-transcribed 7SK gene serves as an internal control since this should be unaffected by the levels of α-amanitin used. When α-amanitin is added before the addition of the radioactive nucleotide, a signal is detected over probe C, indicating that transcripts produced by either pol I or pol III are hybridizing to this probe. Accordingly, these background signals have been subtracted from the signals over probe C for all other run-on assays. Signals are visible over at least some of the probes downstream of the 3′ box as early as 1 min. However, the signals at this time point are too low to be quantitated accurately. Therefore, only the results for the 2, 5 and 10 min time points are depicted graphically in Figure 3B after normalizing to the level of the 7SK signal.

The signals over probes A–Z increase from 1 to 10 min, indicating that transcription is continuing during this time. After 2 min, the signals are quite evenly distributed over probes A–J. Although all the signals have increased ~10-fold from the 2 to 10 min time points (relative to 7SK), the pattern has not changed markedly, apart from a relative increase over probe C. This may suggest that polymerases are stacking over probe C or that some of this signal is due to cross-hybridization to non-U2 gene transcripts. Importantly, there is no obvious increase in the signals over the downstream probes relative to probes A–D with time, and the absolute signal over probe A increases with time. Reinitiation of transcription by pol II is

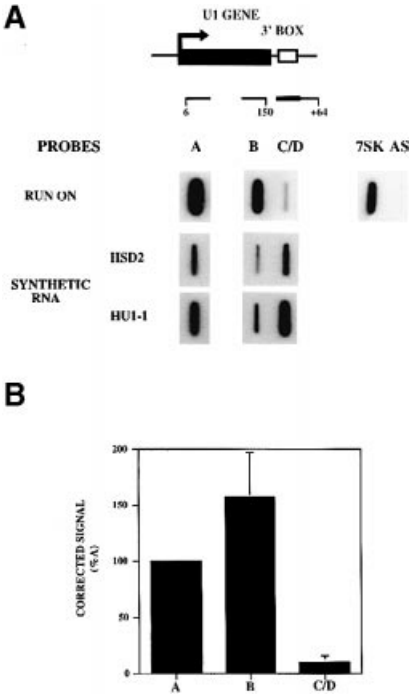

Fig. 4. Termination of transcription appears to occur close to the 3′ box of the U1 genes in our assay system. (**A**) A diagram of the structure of the U1 gene is shown with the relative positions of the probes marked below. The numbers noted next to the probes indicate the start or end of the probes used relative to the site of initiation (A and B) or relative to the end of the 3′ box (C and D). Oligos C and D are complementary to transcripts from the *HSD2* and *HU1-1* U1 alleles, respectively (see text). The results of a run-on analysis and hybridization of synthetically produced RNA to the probes are shown below each probe. (**B**) A graphic representation is shown of the results of the run-on analysis in (A) as a percentage of the signal over probe A after subtraction of α-amanitin-resistant transcription and correction for the hybridization efficiency of each probe.

very inefficient in nuclear run-ons (Bentley and Groudine, 1986), and transcripts of mRNA genes are elongated by ~100 nucleotides in 15 min under our run-on conditions (Dye and Proudfoot, 1999). It is therefore likely that transcription is still running-on over the 47 bp region of probe A within the 10 min of the time course.

These results strongly suggest that polymerases are not artefactually reading through a termination signal under the run-on conditions.

Termination of transcription appears to occur close to the 3′ box of the U1 genes in our assay system

As mentioned above, Kunkel and Pederson (1985) found that transcription of the human U1 genes terminates close to the 3′ box. Our assay conditions are not identical to theirs and they used longer M13 probes to detect run-on transcripts. We obtained essentially identical results to those shown above for the U2 genes using the buffer conditions described by Kunkel and Pederson (1985), although the signals were lower over all probes (data not shown). We have also carried out an analysis of the U1 genes using oligonucleotide probes (Figure 4). To investigate the polymerase density across these genes, we have used two 50mer oligonucleotides complementary to sequences within the coding region (A and B). The sequences downstream from the 3′ box of the U1 genes are not identical in the seven of the 15–30 U1 genes sequenced so far, although they are highly related for up to at least 70 bp past this element (Manser and Gesteland, 1982; Lund and Dahlberg, 1984). We have therefore used a mixture of two longer (80mer) oligonucleotide probes (C and D) complementary to the region from 8 to 88 bp downstream of the coding region of the *HSD2* U1 gene (Manser and Gesteland, 1982) and the *HU1-1* U1 gene (Lund and Dahlberg, 1984) to increase the possibility of detecting transcripts from this region. Oligos C and D include the complement to the 3′ box which spans the sequence from 11 to 24 bp downstream of the coding region.

The hybridization efficiency of the probes A–D was determined using labelled synthetic RNAs (Figure 4A) and the nuclear run-on results were normalized to these values after subtraction of α-amanitin-resistant transcription. The position of each probe relative to the U1 gene and the results of a representative run-on analysis are shown in Figure 4A. The 7SK and AS oligonucleotides have also been included as positive and negative controls, respectively. The profile of polymerase density over the U1 genes determined by three such experiments is shown in Figure 4B as a percentage of the level of signal over probe A.

Transcription appears to be much higher over the coding region of the U1 genes than directly downstream, where it has dropped to 10% of the value for probe A. This suggests that termination of transcription is occuring close to the 3′ box in these genes. The signal detected over probes C/D may indicate that transcription continues just into this region or that some transcripts corresponding to the whole length of the probes are made. Alternatively, the probes we have used may only detect a small proportion of the transcripts produced by the U1 genes due to sequence divergence downstream of the 3′ box. However, probes C and D together should detect transcripts from at least the seven sequenced U1 genes (see Manser and Gesteland, 1982; Lund and Dahlberg, 1984). We therefore think it likely that our results are due to termination of transcription of the U1 genes close to the 3′ box in our assay system and that they confirm the findings of Kunkel and Pederson (1985).

These results are consistent with the notion that the 3′ box is a termination signal. However, the 3′ box of the U2 gene would, therefore, be a very inefficient transcription termination signal in comparison.

In vivo footprinting detects DNA–protein interactions downstream of the 3′ box of the U1 gene

The function of the 3′ box requires that transcription is initiated specifically from a class II snRNA promoter (Hernandez and Weiner, 1986; Neuman de Vegvar *et al*., 1986), indicating a strong link between initiation of transcription and RNA processing, as has recently been shown for processing of mRNA (reviewed by Corden and Patturajan, 1997; Neugebauer and Roth, 1997; Steinmetz, 1997).

In our studies of the DNA–protein interactions required for initiation of transcription of snRNA genes, we have carried out an extensive *in vivo* footprinting analysis of the human U2 genes (D.C.Boyd, I.Greger and S.Murphy,

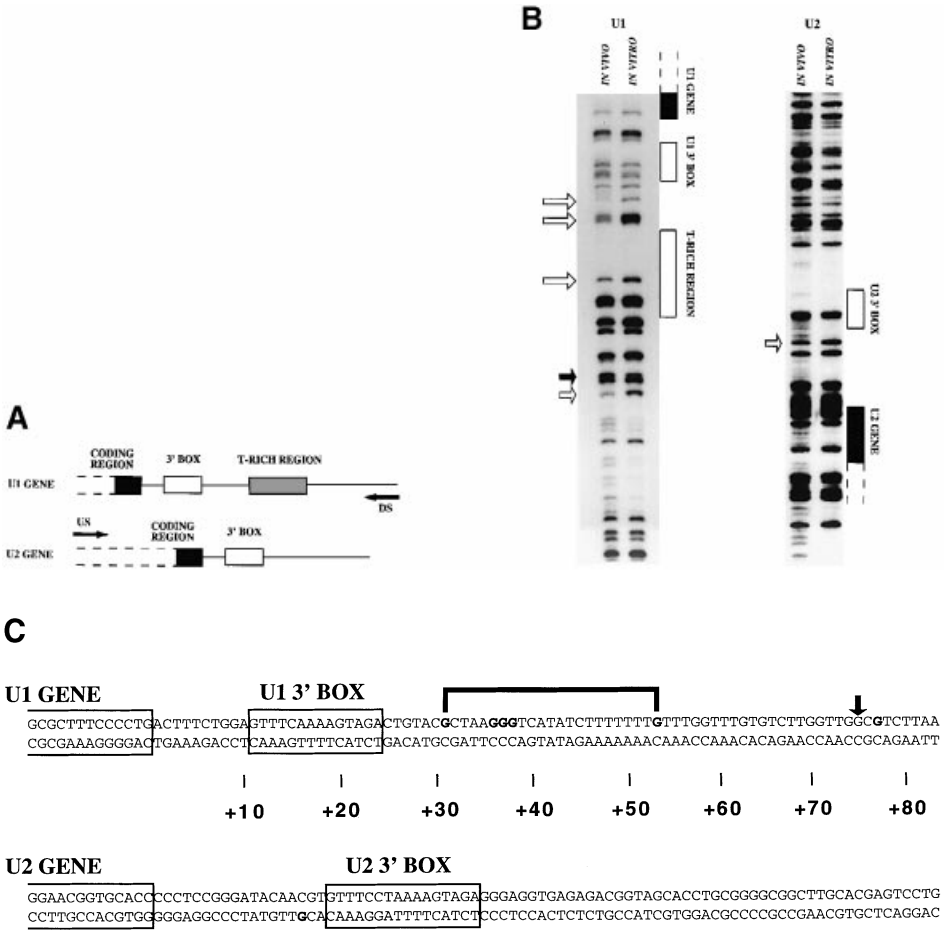

Fig. 5. *In vivo* footprinting detects DNA–protein interactions downstream of the 3′ box of the U1 gene. (**A**) A diagram of the structure of the U1 and U2 genes is shown with the relative positions of the *in vivo* footprinting primers indicated by black arrows. (**B**) The results of *in vivo* footprinting over the region of the 3′ box of the U1 and U2 genes is shown. The white arrows indicate a reduction in the relative intensity of bands (Gs) in the *in vivo* treated DNA compared with *in vitro* modified DNA, and the size of the arrow indicates the magnitude of the reduction. The black arrow indicates a moderate increase in the intensity of a band in the *in vivo* treated DNA. (**C**) The sequence of the 3′ box region of the *HU1-1* U1 gene (Lund and Dahlberg, 1984) and the pTP18 U2 genes (Pavelitz *et al*., 1995) are shown. The numbers under the U1 sequence indicate the distance from the 3′ box. G residues that were protected from methylation by DMS *in vivo* are indicated in bold. The region downstream of the U1 3′ box that is footprinted is indicated by a bracket and a G residue that is hypersensitive after DMS treatment *in vivo* is marked by a black arrow.

unpublished results). Since footprints are easily detected over all known promoter elements of these genes, it is likely that a large number of the genes are active at any one time and that DNA–protein interactions required for termination of transcription may also be detected. We have therefore analysed the *in vivo* DNA–protein interactions over the 3′ box region of the U1 and U2 genes using DMS footprinting.

The results of this analysis are shown in Figure 5A and B and summarized graphically in Figure 5C. For the U1 gene, the upper strand of *HU1-1* (Lund and Dahlberg, 1984) has been analysed and for the U2 gene the lower strand has been analysed (see Figure 5A). No methylation protection is detected over the 3′ box of either the U1 or U2 genes, although a G residue just upstream from the U2 3′ box was partially protected from methylation (marked in Figure 5B and C). However, a region of very clear protection from methylation is apparent immediately downstream from the 3′ box of the U1 gene, close to a T-rich stretch of sequence (Figure 5B and C), suggesting that a protein is binding here *in vivo*. Importantly, the central 11 nucleotides of the footprint are highly conserved in all of the U1 genes sequenced so far (see Manser and Gesteland, 1982; Lund and Dahlberg, 1984).

Although these results do not rule out that protein(s) bind to the 3′ box and directly mediate termination, they are consistent with the hypothesis that termination of transcription occurs downstream of this element in both the U1 and U2 genes. In addition, they raise the possibility that protein(s) binding just downstream of the 3′ box mediate termination of transcription of the U1 gene.

The 3′ box is required for high levels of correctly processed U2 RNA

To test the function of the 3′ box and downstream sequences in processing and termination, we constructed a series of U2 genes, marked by the addition of a linker to the coding region, where the 3′ box is either present or absent or has been replaced by U1 sequences (Figure 6A). These constructs were transfected into HeLa cells, and the 5′ and 3′ ends of the transcribed RNAs mapped by S1 and RNase protection analysis (Figure 6B and C). The gene for the pol III-dependent VAI RNA was included as a co-transfection control. The probes used for these

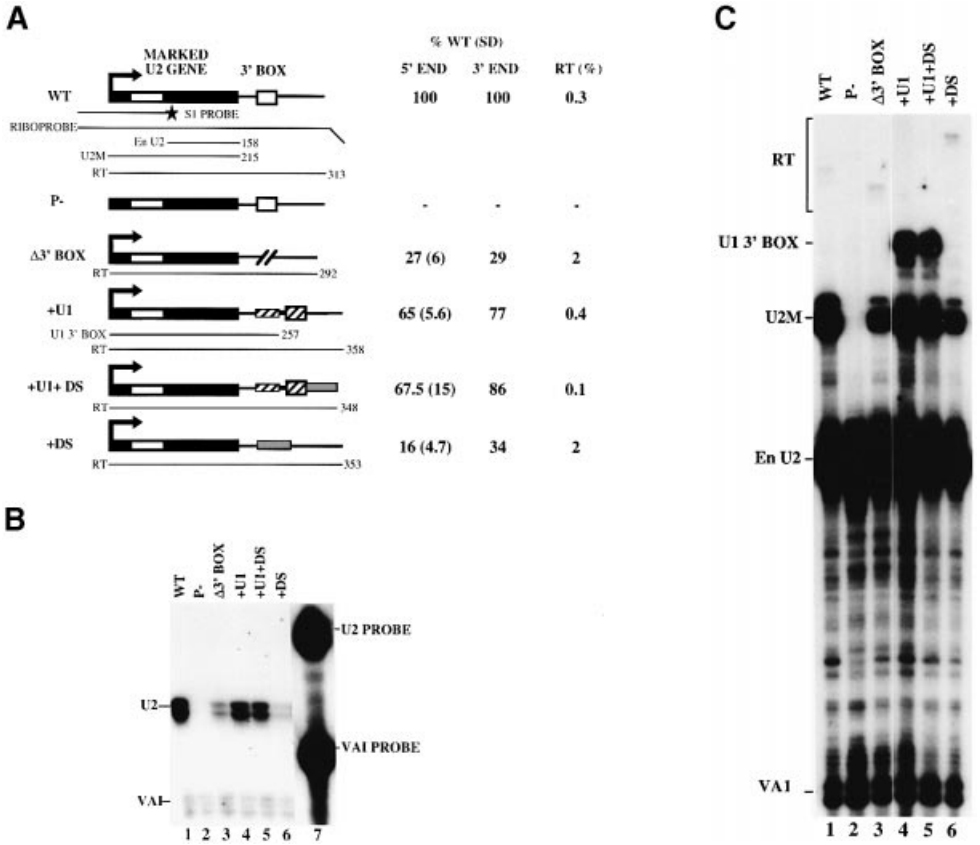

Fig. 6. The 3′ box is required for high levels of correctly processed U2 RNA. (**A**) The structure of marked U2 genes with changes to the 3′ box region is represented. WT is identical to the endogenous U2 gene apart from the insertion of a linker in the coding region. P– is transcriptionally inactive due to a mutation in the promoter. Δ3′ BOX has had the 3′ box deleted. In +U1, the 3′ box is replaced by the last 26 bp of the U1 gene (small cross-hatched box) and the 3′ box (large cross-hatched box), and +U1+DS contains the 63 bp just downstream of the U1 3′ box in addition (grey box). In +DS, the U2 3′ box has been replaced by the sequences downstream from the U1 3′ box. The relative positions of the S1 probe and riboprobes is noted below the WT structure. The sizes of the specific U2 RNase protection products for the constructs are noted below the WT structure. The size of the additional product from +U1 and +DS is indicated below the +U1 structure. The size of the readthrough products is indicated below each structure. The results of quantitation of S1 (mean and standard deviation) and RNase protection analysis (average of two experiments) are noted on the right of the figure. (**B**) The results of S1 analysis of RNA transcribed from the constructs in (A) is shown. The construct is noted above each lane, the positions of the S1 products is noted at the left and the position of the probes is noted at the right. (**C**) The results of RNase protection analysis of RNA transcribed from the constructs in (A) is shown. The construct is noted above each lane. The positions of the protected products are noted on the left.

analyses are marked on Figure 6A, and the quantitation of the results is shown to the right of the figure. Since the 5′ S1 probe is complementary to a common region of the constructs, the relative amount of steady-state RNA can be readily assessed (Figure 6B). A high level of transcription is detected from the 'wild-type' marked U2 gene (WT, lane 1), as seen previously (Murphy, 1997), and mutation of the essential PSE element in the promoter (P–, lane 2) abolishes transcription as expected. Mutation of the 3′ box causes a significant reduction in the amount of steady-state RNA (Δ3′ BOX, lane 3, 27% of WT). This is in agreement with previous studies on human snRNA genes showing that the 3′ box is essential for the accumulation of high steady-state levels of transcripts (Hernandez, 1985; Yuo *et al*., 1985; Neuman de Vegvar *et al*., 1986; Ach and Weiner, 1987; reviewed in Hernandez, 1992). Addition of the 3′ box of the U1 gene to the Δ3′ BOX construct increases the amount of steady-state RNA (+U1, lane 4, 65%), confirming that the 3′ boxes are largely interchangeable. Replacement of the U2 3′ box and downstream region with the equivalent region of the U1 gene has essentially the same effect as replacing the 3′ box alone (+U1+DS, lane 5, 67.5%), indicating that this downstream, *in vivo*-footprinted region has little effect on the steady-state RNA. Hernandez (1985) also saw little effect of this downstream region on the formation of 3′ ends of transcripts from a U1 promoter. Addition of the U1 downstream region alone to the Δ3′ BOX template only further decreased the level of steady-state RNA detected (+DS, lane 6, 16%).

The RNase protection analysis shown in Figure 6C demonstrates the effect of the changes on the 3′ ends of transcripts. In this analysis, the endogenous U2 is complementary to a region in all the probes used (see Figure 6A) and a protection product of 158 nucleotides (EnU2) is therefore present in all lanes. The VAI RNA gives an 89 nucleotide product that serves as an internal transfection control. The major product of the WT construct is a band of the expected size for mature marked U2 RNA (WT, lane 1, U2M). There are also distinct products several nucleotides longer that may correspond to the 3′-extended precursors of U2 RNA previously described (Wieben *et al*., 1985; Yuo *et al*., 1985). The same pattern is evident for the endogenous U2 RNA also

detected by this analysis (EnU2). The much longer product visible in lane 1 corresponds to small amounts of readthrough transcripts mismatching with the end of the probe (RT, 0.3% of total transcripts) (see Figure 6A). Low levels of readthrough transcripts from the endogenous genes are also detected in longer exposures of RNase protection analyses (data not shown). Apart from the endogenous U2 products, none of these bands is detected when the promoter of the U2 gene is disabled (P–, lane 2). When the 3′ box is deleted, the amount of mature marked U2 and the potential precursors decreases (Δ3′ BOX, lane 3). However, the readthrough transcripts increase to 2% of the total. In the +U1 construct, the U2 3′ box has been replaced by the last 26 bp of the U1 gene followed by the 3′ box (see Materials and methods). Thus, the 3′ end directed by the U1 sequences should be distinct from that of the WT construct, and products of the right size (257 nucleotides) are detected for the +U1 and +U1+DS constructs (lanes 4 and 5, U1 3′ BOX). Again, potential precursors extended by a few nucleotides at the 3′ end are detected. Perhaps surprisingly, an approximately equal level of products equivalent to the mature U2 3′ end and associated precursors is also apparent (see below). Approximately the same level of readthrough transcripts is detected for +U1 as for the WT construct (lane 4, 0.4%). However, the level of readthrough appears to have decreased when the region downstream of the U1 3′ box is present (+U1+DS, lane 5, 0.1%). Addition of this downstream region alone to the Δ3′ box construct has very little effect on any of the products (+DS, lane 6).

The results of this analysis confirm that the U2 3′ box is necessary for the accumulation of high levels of properly processed transcripts from the marked U2 gene. However, it appears that proper 3′ end formation is still occurring, albeit at a reduced level, in the complete absence of the 3′ box. This has also been noted in previous studies (Neuman de Vegvar *et al*., 1986; Ach and Weiner, 1987) and suggests that this process can occur by more than one pathway. The RNA produced from these constructs should interact with the appropriate proteins to make U2 snRNP, including the Sm-binding proteins. It is possible, therefore, that longer transcripts, produced in the absence of the U2 3′ box, with or without the addition of the U1 3′ box downstream, can be trimmed after complexing with proteins in the cytoplasm. However, the pattern of bands is identical in the absence and presence of the 3′ box, suggesting that these 'precursors' are generated in the same way in both cases. Thus, the slightly 3′-extended transcripts noted above U2M in lane 1 (WT construct) may not represent specific precursors formed by a 3′ box-dependent mechanism. Alternatively, other sequences within the gene may be able to direct a low level of proper 3′ end formation in the absence of the 3′ box.

Since the transcripts in this analysis are variable in sequence, any differences noted may simply reflect changes in RNA stability. However, the drop in detectable readthrough products from the +U1+DS construct is consistent with the possibility that the downstream *in vivo*-footprinted region of the U1 gene is capable of terminating transcription.

The region downstream from the 3′ box of the U1 contains a transcription terminator

In order to determine unequivocally whether the 3′ box is causing termination directly, we have carried out

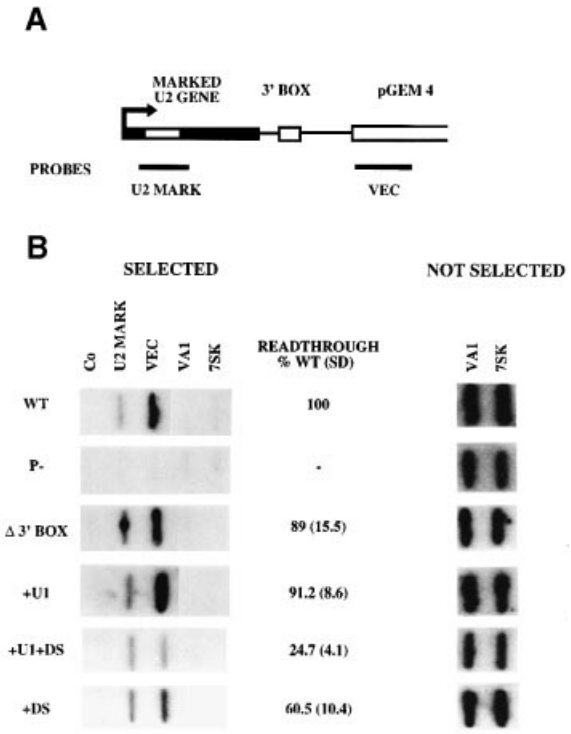

Fig. 7. The region downstream from the 3′ box of the U1 contains a transcription terminator. (**A**) The structure of the marked U2 gene construct is shown and the relative positions of the probes used for the run-on analysis are indicated. (**B**) The results of run-on analysis of the constructs in Figure 5A are shown. The hybridizations on the left of the figure were carried out with RNA that had been hybrid-selected using a probe complementary to sequences in the 'coding' RNA from the marked template (see text). The hybridizations on the right were carried out with the RNA in the supernatant from the selection that contains any RNA that has not hybridized to the probe. The probes present on the filters are noted above each slot and the constructs transfected are noted on the left. The quantitation of the relative levels of hybridization over the U2 MARK and VEC probes is expressed as a percentage of the ratio for the WT construct.

run-on analysis on the transfected U2 templates (Figure 7). The *VAI* gene was again included as a transfection control and the probe for 7SK was used to control the efficiency of the run-on assay. U2 MARK is a 50mer oligonucleotide complementary to the linker in the marked U2 gene plus some flanking sequence (see Materials and methods) used to detect transcription of the coding sequence. The probe VEC is an 80mer oligonucleotide complementary to the region of the vector downstream from the U2 3′ box which detects readthrough transcription (Figure 7A). Co is an 80mer oligonucleotide complementary to a region of the vector upstream from the U2 promoter and serves as a background control. After carrying out the run-on reaction with nuclei from transfected cells, RNA was hybrid-selected using a probe complementary to part of the coding region of the marked U2 gene (see Materials and methods). The selected RNA was then hydrolysed and hybridized to nitrocellulose filters (Figure 7B). This procedure ensures that signals are due to RNA contiguous with the coding region and greatly reduces the level of background, allowing an accurate quantitation of the low signals over the short U2 MARK probe.

RNA from cells transfected with the WT construct hybridizes to both the U2 MARK and VEC probes,

indicating that transcription is continuing beyond the 3′ box in the transfected gene as it does in the endogenous U2 genes. Importantly, no signal is detected over any U2 template probe when the PSE is mutated (Figure 7B, P–). Neither VAI nor 7SK transcripts have been selected by the probe and remain in the supernatant after the selection procedure, indicating that the selection is specific. Deletion of the U2 3′ box has little effect on the level of readthrough (Δ3′ BOX, 89% of the level seen with WT). Introduction of the 3′ box of the U1 gene similarly has little effect (+U1, 91.2%). However, the presence of both the U1 3′ box and the immediate downstream region significantly reduces the level of readthrough (+U1+DS, 24.7%). The addition of the U1 downstream region alone appears to have a less pronounced effect on the level of readthrough (+DS, 60.5%). The levels of VAI transcripts do not appear to vary drastically relative to the coding region transcripts of the different transfected genes, suggesting that changes in the levels of steady-state RNA (Figure 6) are due mainly to post-transcriptional effects.

These results lend further support to the hypothesis that the *in vivo*-footprinted element located immediately downstream of the 3′ box of the U1 genes is a transcriptional terminator. This would satisfactorily explain the difference in the run-on profiles of the human U1 and U2 genes. The U1 and U2 3′ boxes alone appear to have little effect on the level of readthrough transcription, indicating that these are not efficient termination signals by themselves. However, the putative U1 terminator also does not appear to function efficiently in the absence of the 3′ box, suggesting that both elements are necessary for termination of transcription.

Discussion

Previous studies on snRNA gene transcription have not resolved whether the 3′ box functions in the nascent RNA as a processing element or as a DNA element to terminate transcription. The 3′ box is necessary for efficient 3′ end formation of snRNA gene transcripts (Hernandez, 1985; Yuo *et al*., 1985; Ciliberto *et al*., 1986; Neuman de Vegvar *et al*., 1986; Ach and Weiner, 1987; reviewed in Hernandez, 1992), and the function of this element cannot be separated readily from the process of transcription. Notably, transcripts containing a copy of the 3′ box do not undergo processing when reinjected into *Xenopus* oocytes (Ciliberto *et al*., 1986) and although U1 transcripts that include the 3′ box are detected in HeLa cells, these do not seem to be processing substrates (Lobo and Marzluff, 1987). In addition, recognition of the U1 3′ box *in vitro* appears to be co-transcriptional and only occurs if transcription is initiated from a class II snRNA promoter (Gunderson *et al*., 1990). Although these findings do not rule out that the 3′ box is a processing signal, they support the notion that this element is a transcription terminator. The demonstration by Kunkel and Pederson (1985) that transcription terminates very close to the 3′ box in the U1 genes provided the most compelling argument to date that the 3′ box is a termination element. However, the results presented here show that the 3′ box is not sufficient for efficient termination of transcription. Instead, we implicate a novel element immediately downstream of the 3′ box in termination of transcription of the U1 gene. Termination may be effected directly by proteins binding to the sequence within the *in vivo*-footprinted region, perhaps in conjunction with a high proportion of T residues. Interestingly, the results of our nuclear run-on analysis indicate that the human U2 genes do not have an efficient termination element within the 250 bp downstream of the 3′ box. Our results also suggest that termination is coupled to recognition of the 3′ box.

Why does the U1 gene have a terminator close to the 3′ box?

It is not clear why termination of transcription of the U1 genes should occur very close to the 3′ box while transcription of the U2 genes continues further downstream. The U1 gene repeat unit is >45 kb long and repeated 15–30 times on chromosome 1 (Lindgren *et al*., 1985b). This is an extraordinarily large repeat unit considering that the size of the RNA-coding region of the U1 gene is <200 bp. Unlike the U2 repeat units which are arranged tandemly, the U1 genes are loosely clustered and organized in an irregular and highly polymorphic manner (Manser and Gesteland, 1982; Lund and Dahlberg, 1984; Bernstein *et al*., 1985). In addition, there are multiple genes present in the *RNU1* locus. These genes include tRNA genes (Van der Drift *et al*., 1994, 1995) and several putative neuroblastoma suppressor genes including human Kruppel-related 3 (*HKR3*) (Maris *et al*., 1997) and human Elk-related kinase (*ERK*) (Saito *et al*., 1995). Since not all the genes in this region have been identified, or their precise locations determined, it is unclear whether any of these genes lie close to the U1 genes. Perhaps termination of transcription of the U1 genes is more critical than for the U2 genes, in order to prevent polymerases reading into downstream genes and causing inhibition of initiation by transcriptional interference [see Greger *et al*. (1998) for a discussion of this phenomenon].

How might initiation, 3′ box function and termination be linked?

It is possible that only a fraction of the nascent transcripts from the human U2 genes are terminated at the 3′ box and processed properly. However, this seems unlikely as these genes encode very abundant RNAs. In common with Yuo *et al*. (1985) and Neuman de Vegvar *et al*. (1986), we favour the hypothesis that the 3′ box is an RNA processing element present in the nascent RNA and that the 3–16 nucleotide extended precursors detected *in vivo* are produced directly by an RNA processing event mediated by the 3′ box. Yuo *et al*. (1985) have speculated further that such a processing reaction may involve snRNAs in analogy to 3′ end processing of histone mRNAs (see below).

The 3′ ends of all other pol II transcripts (mRNAs) are produced by cleavage of nascent RNA. The replication-activated histone mRNAs are produced from precursors by a cleavage event that requires base pairing between the nascent RNA and the U7 snRNA (reviewed by Marzluff, 1992). In the case of the remaining mRNAs, cleavage of the nascent RNA and subsequent polyadenylation of the resulting 3′ end are directed by the polyadenylation signal in the transcribed RNA (reviewed by Colgan and Manley, 1997). Processing at the poly(A) site is also linked intimately to both initiation

and termination of transcription. Factors required for polyadenylation associate with the C-terminal domain (CTD) of pol II before the poly(A) signal is transcribed, possibly at initiation (Dantonel *et al*., 1997), and removal of the CTD dramatically decreases the efficiency of polyadenylation *in vivo* (McCracken *et al*., 1997). In addition, termination of transcription of vertebrate mRNA genes requires that splicing of the terminal intron and polyadenylation take place, even though termination can occur hundreds of base pairs downstream from the coding region (Dye and Proudfoot, 1999, and references therein). Exactly how termination occurs in these genes is not well understood, but Birse *et al*. (1998) have shown recently that the function of cleavage factors in the polyadenylation apparatus is required for termination of transcription of mRNA genes in *Saccharomyces cerevisiae*. Processing of the RNA at the poly(A) site is likely to influence directly the association of processing factors with the CTD of the transcribing polymerase. Thus, transcribing a poly(A) signal may convert the polymerase to a more easily terminated form as suggested by McCracken *et al*. (1997) and Dye and Proudfoot (1999). In the case of the closely spaced genes encoding the mammalian complement proteins, C2 and factor B, however, it has been shown that binding sites for the protein Maz are required in addition to a functional poly(A) site for efficient termination to occur (Ashfield *et al*., 1994). Thus, a combination of processing and binding of factors to elements in the DNA can work together to terminate pol II relatively close to the poly(A) site.

Efficient recognition of the 3′ box *in vivo* and *in vitro* requires that transcription is initiated from a class II snRNA gene promoter, indicating that the pre-initiation complex formed on these genes influences the downstream RNA processing event(s) (Hernandez and Weiner, 1986; Neuman de Vegvar *et al*., 1986; Gunderson *et al*., 1990). Thus, the initial 3′ end processing event of transcripts from both mRNA and snRNA genes appears to be influenced directly by events at initiation. The mechanism of co-ordinate promoter/3′ box-directed processing of snRNA gene transcripts may be analogous to the mechanism of polymerase-assisted 3′ end cleavage and polyadenylation of mRNA genes. As mentioned above, the structure of the promoters of class II snRNA genes is different from that of mRNA genes. The essential PSE element is both necessary and sufficient for initiation and 3′ box recognition *in vivo* (Hernandez and Lucito, 1988; Neuman de Vegvar and Dahlberg, 1989; Parry *et al*., 1989). The PSE is recognized by a multisubunit factor known variously as PTF, SNAPc and PBP (Henry *et al*., 1998, and references therein), which is critical for the formation of a specific pre-initiation complex containing TBP and TFIIB (Bernues *et al*., 1993; Sadowski *et al*., 1993). This specialized complex may directly recruit a specialized polymerase associated with processing factors. Alternatively, the polymerase may become associated with specific processing factors after recruitment, as appears to happen in the interaction of the CTD of pol II with splicing and polyadenylation factors. An obvious scenario is one where snRNA processing factors rather than mRNA processing factors interact with pol II at initiation of transcription of snRNA genes, as already suggested by Neuman de Vegvar *et al*. (1986). This would account for the finding that polyadenylation signals are not recognized efficiently in transcripts initiated from the promoters of the U1 or U2 genes (Neuman de Vegvar *et al*., 1986; Dahlberg and Schenborn, 1988; Hernandez and Lucito, 1988; Lobo and Hernandez, 1989). However, recognition of poly(A) signals can be uncoupled readily from transcription, whereas recognition of the 3′ box cannot (Lobo and Marzluff, 1987; Ciliberto *et al*., 1989; Gunderson *et al*., 1990), as noted above. This suggests that a promoter-dependent activity is also essential for any processing event directed by the 3′ box.

Termination of transcription of both mRNA genes and snRNA genes may also occur by a common mechanism involving cleavage of the transcript/release of processing factors from the polymerase. Termination of transcription of the U2 gene would, then, resemble the process in those mRNA genes where termination occurs hundreds of base pairs downstream from the poly(A) site. In the U1 genes instead, factors bound to the DNA just downstream of the processing element terminate the 'destabilized' polymerase much sooner.

Materials and methods

Nuclear run-ons and hybrid selection

The protocol of Ashe *et al*. (1997) was used with the following changes: ~50 μl of nuclei were collected from two 120 mm dishes of HeLa cells for each reaction and 7 μl of [α-^{32}P]UTP (3000 Ci/mmol) was added to each reaction mix. A 1 μg aliquot of each 47–50mer oligonucleotide or 1.6 μg of each 80mer oligonucleotide was denatured and fixed to the nitrocellulose filter after slot-blotting. After hybridization, filters were washed three times in 1.0× SSPE/0.1% (w/v) SDS at 37°C (sometimes with 1 μg/ml RNase A added) and exposed to Kodak AR film at –70°C or to Molecular Dynamics Phosphor storage screens for quantitation in a PhosphorImager.

For run-on analysis of the endogenous U2 genes, the oligos are complementary to the U2 gene sequence from pTP18 (Pavelitz *et al*., 1995; DDBJ/EMBL/GenBank accession No. U57614) taking the first base pair of the U2 coding sequence as 1: US, –50 to –1; A, 1–47; B, 48–94; C, 95–141; D, 142–188; E, 189–238; F, 239–288; G, 289–338; H, 339–388; I, 399–438; J, 439–489; Z, 789–838. AS is complementary to A. The 7SK oligonucloeotide is complementary to sequences 61–110 of the coding region for 7SK (Murphy *et al*., 1986).

For run-on analysis of the U1 genes, the oligos A, B and C are complementary to nucleotides 6–55, 101–150 and 171–250, respectively, of *HSD2* (Manser and Gesteland, 1982) taking the first base of the U1 coding region as 1. Oligo D is complementary to nucleotides 171–250 of *HU1-1* (Lund and Dahlberg, 1984; DDBJ/EMBL/GenBank accession No. J00318).

For run-on analysis of transfected genes, the nuclei were collected from two 120 mm dishes of HeLa cells 48 h after transfection with 12.5 μg of the test construct and 1.25 μg of VAI using Lipofectamine (Gibco-BRL) according to the manufacturer's instructions. U2 MARK is 5′-GAT- ACTACACTTGATCCTCTAGAGCTGGTACCACTAGAGG-ATCTTAGCCA-3′. VEC and Co are complementary to bp 61–140 and 2792 to +2871 of pGEM 4, respectively, taking the first base pair of the SP6-transcribed region as 1. The VAI oligonucleotide is complementary to bp 1–80 of the VAI-coding region. Synthetic RNAs were made by T7 or SP6 from sequences cloned into pGEM vectors. Hybrid selection probes were made in the same way with the inclusion of biotin-16-UTP (Boehringer Mannheim) as described by Dye and Proudfoot (1999). Hybrid selection was carried out as described by Dye and Proudfoot (1999).

Constructs

The marked U2 gene is described by Murphy (1997) and extends from the *Stu*I site at –556 of dl-556 (Ares *et al*., 1985) to the *Nae*I site at +93 downstream of the end of the RNA coding region. This was cloned into pGEM4 using the *Eco*RI and *Hin*cII sites in the polylinker. P– was made by PCR to replace the sequence between –60 and –49 by TCCCACGGCCGT. The Δ3′ BOX construct was made by PCR to

replace the 3′ box from +19 to +41 with a C residue to create a unique *Mlu*I site. The U1 sequence from 138 to 187 and the sequence from 192 to 250 of pHU1-1 (Lund and Dahlberg, 1984) downstream of the start of transcription were cloned into this site to create +U1 and +DS, respectively. For +U1+DS, *Mlu*I–*Nae*I was replaced with sequences from 138 to 250 of *HU1-1*. The pGEM 4-based template for the hybrid selection probe for endogenous U2 (Figure 2) was made by PCR from pTP18 (Pavelitz *et al*., 1995) using oligos AS and D. For hybrid selection of marked U2 transcripts, a fragment containing the sequence from the *Dde*I site at bp 20 of the coding region to the end of oligonucleotide D was prepared by PCR of the marked U2 template and cloned into pGEM3. Each U2 construct was recloned into the *Eco*RI site of pGEM4 to place the polylinker region between the T7 promoter and the 3′ end of the U2 sequence.

Steady-state RNA analysis

For each 90 cm dish of HeLa cells, a 5 μg aliquot of each U2 construct was co-transfected with 500 ng of *VAI* using Lipofectamine (Gibco-BRL) as recommended by the manufacturers, and cytoplasmic RNA was collected as described by Whitelaw *et al*. (1989). RNase protection was carried out as described by Eggemont and Proudfoot (1993). Riboprobes for analysis of U2 transcripts were made by T7 after digestion with *Fok*I that cuts 42 bp upstream from the start of the U2 coding region. The VAI riboprobe was made by SP6 after digesting the *VAI* gene in pGEM4 with *Bam*HI. S1 analysis was carried out as described by Murphy (1997) using the oligonucleotides described therein. The products of S1 analysis are 45 and 63 nucleotides for the VAI and U2 RNAs, respectively.

In vivo footprinting

Cells or purified naked DNA were treated first with DMS (Aldrich) and then with piperidine (Aldrich) to prepare *in vivo* or *in vitro* DNA samples, respectively, as described by Greger *et al*. (1998). Ligation-mediated PCR was carried out on these samples using the following nested primers for U1: U1.DS1, 5′-CTCGCTTTTTCTCCTATGGC-3′; U1.DS2, 5′-CAGGCGACATGTTACTTCCTATTCCGC-3′; U1.DS3, 5′-GGCGACATGTTACTTCCTATTCCGCAGCCCTC-3′; and for U2: U2.US1, 5′-GGATTTTTGGAGCAGGGAGA-3′; U2.US2, 5′-AATAGGAGCTTGCTCCGTCCACTCC-3′; U2.US3, 5′-AGGAGCTTGCTCCGTCCACTCCACGCATC-3′.

Acknowledgements

We would like to thank Andre Furger and David Bentley for helpful suggestions, and Andre Furger and Alexandra Moreira for critical reading of the manuscript. We would especially like to thank Charlie Birse for encouraging us to try using oligonucleotide probes for the run-ons and Ingo Greger for setting up *in vivo* footprinting in the laboratory. P.C. was supported by a BBSRC graduate studentship, M.D. was supported by a Glaxo graduate studentship, N.J.P. is supported by Wellcome Programme Grant No. 032743, and D.C.B. and S.M. are supported by MRC Senior Fellowship No. G117/309.

References

Ach,R.A. and Weiner,A.M. (1987) The highly conserved U small nuclear RNA 3′-end formation signal is quite tolerant to mutation. *Mol. Cell. Biol.*, **7**, 2070–2079.

Ares,M.,Jr, Mangin,M. and Weiner,A.M. (1985) Orientation-dependent transcriptional activator upstream of a human U2 snRNA gene. *Mol. Cell. Biol.*, **5**, 1560–1570.

Ashe,H.L., Monks,J., Wijgerde,M., Fraser,P. and Proudfoot,N.J. (1997) Intergenic transcripts and transinduction of the human β-globin locus. *Genes Dev.*, **11**, 2494–2509.

Ashfield,R., Patel,A.J., Bossone,S.A., Brown,H., Campbell,R.D., Marcu,K.B. and Proudfoot,N.J. (1994) MAZ-dependent termination between closely spaced human complement genes. *EMBO J.*, **13**, 5656–5667.

Bentley,D.L. and Groudine,M. (1986) A block to elongation is largely responsible for decreased transcription of c-*myc* in differentiated HL60 cells. *Nature*, **321**, 702–706.

Bernstein,L.B., Manser,T. and Weiner,A.M. (1985) Human U1 small nuclear RNA genes: extensive conservation of flanking sequences suggests cycles of gene amplification and transportation. *Mol. Cell. Biol.*, **5**, 2159–2171.

Bernues,J., Simmen,K.A., Lewis,J.D., Gunderson,S.I., Moncollin,M., Egly,J.-M. and Mattaj,I.W. (1993) Common and unique transcription factor requirements of human U1 and U6 snRNA genes. *EMBO J.*, **12**, 3575–3585.

Birse,C.E., Minvielle-Sebastia,L., Lee,B.A., Keller,W. and Proudfoot,N.J. (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science*, **280**, 298–301.

Chanfreau,G., Abou Elela,S., Ares,M. and Guthrie,C. (1997) Alternative 3′-end processing of U5 snRNA by RNase III. *Genes Dev.*, **11**, 2741–2751.

Ciliberto,G., Dathan,N., Frank,R., Philipson,L. and Mattaj,I.W. (1986) Formation of the 3′ end on U snRNAs requires at least three sequence elements. *EMBO J.*, **5**, 2931–2937.

Colgan,D.F. and Manley,J.L. (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev.*, **11**, 2755–2766.

Corden,J.L. and Patturajan,M. (1997) A CTD function linking transcription to splicing. *Trends Biochem. Sci.*, **22**, 413–416.

Dahlberg,J.E. and Schenborn,E.T. (1988) The human U1 snRNA promoter and enhancer do not direct synthesis of messenger RNA. *Nucleic Acids Res.*, **16**, 5827–5840.

Dantonel,J.C., Murthy,K.G.K., Manley,J.L. and Tora,L. (1997) Transcription factor TFIID recruits factor CPSF for formation of the 3′ end of mRNA. *Nature*, **389**, 399–402.

Dye,M.J. and Proudfoot,N.J. (1999) Terminal exon definition occurs co-transcriptionally and promotes termination of RNA polymerase II. *Mol. Cell*, **3**, 371–378.

Eggermont,J. and Proudfoot,N.J. (1993) Poly(A) signals and transcriptional pause sites combine to prevent interference between RNA polymerase promoters. *EMBO J.*, **12**, 2539–2548.

Elela,S.A. and Ares,J.M. (1998) Depletion of yeast RNase III blocks correct U2 3′ end formation and results in polyadenylated but functional U2 RNA. *EMBO J.*, **17**, 3738–3746.

Greger,I.H., Demarchi,F., Giacca,M. and Proudfoot,N.J. (1998) Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Res.*, **26**, 1294–1300.

Gunderson,S.I., Knuth,M.W. and Burgess,R.R. (1990) The human U1 snRNA promoter correctly initiates transcription *in vitro* and is activated by PSE1. *Genes Dev.*, **4**, 2048–2060.

Hammarstrom,K., Westin,G., Bark,C., Zabielski,J. and Pettersson,U. (1984) Genes and pseudogenes for human U2 RNA. Implications for the mechanism of pseudogene formation. *J. Mol. Biol.*, **179**, 157–169.

Henry,R.W., Mittal.,V., Ma.,B., Kobayashi,R. and Hernandez,N. (1998) SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes Dev.*, **12**, 2664–2672.

Hernandez,N. (1985) Formation of the 3′ end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J.*, **4**, 1827–1837.

Hernandez,N. (1992) Transcription of vertebrate snRNA genes and related genes. In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 1, pp. 281–313.

Hernandez,N. and Lucito,R. (1988) Elements required for transcription initiation of the human U2 snRNA coincide with elements required for snRNA 3′ end formation. *EMBO J.*, **7**, 3125–3134.

Hernandez,N. and Weiner,A.M. (1986) Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. *Cell*, **47**, 249–258.

Huang,Q., Jacobson,M.R. and Pederson,T. (1997) 3′ Processing of human pre-U2 small nuclear RNA: a base-pairing interaction between the 3′ extension of the precursor and an internal region. *Mol. Cell. Biol.*, **17**, 7178–7185.

Kuhn,A. and Grummt,I. (1989) 3′-End formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. *Genes Dev.*, **3**, 224–231.

Kunkel,G.R. and Pederson,T. (1985) Transcription boundaries of U1 small nuclear RNA. *Mol. Cell. Biol.*, **5**, 2332–2340.

Liao,D. and Weiner,A.M. (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the *RNU2* locus) does not prevent rapid diversification of the (CT)n·(GA)n microsatellite embedded within the U2 repeat unit. *Genomics*, **30**, 583–593.

Lindgren,V., Ares,M., Weiner,A.M. and Franke,U. (1985a) Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature*, **314**, 115–116.

Lindgren,V., Bernstein,L.B., Weiner,A.M. and Franke,U. (1985b) Human U1 small nuclear RNA pseudogenes do not map to the site of the U1 genes in 1p36 but are clustered in 1q12–q22. *Mol. Cell. Biol.*, **5**, 2172–2180.

Lobo,S.M. and Marzluff,W.F. (1987) Synthesis of U1 RNA in isolated mouse cell nuclei: initiation and 3′ end formation. *Mol. Cell. Biol.*, **7**, 4290–4296.

Lobo,S.M. and Hernandez,N. (1989) A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell*, **58**, 55–67.

Lobo,S.M. and Hernandez,N. (1994) Transcription of snRNA genes by RNA polymerase II and III. In Conaway,R.C. and Conaway,J.W. (eds), *Transcription, Mechanisms and Regulation*. Raven Press, New York, NY, Vol. 3, pp. 127–159.

Lund,E. and Dahlberg,J.E. (1984) True genes for human U1 small nuclear RNA: copy number, polymorphism and methylation. *J. Biol. Chem.*, **259**, 2013–2021.

Manser,T. and Gesteland,R.F. (1982) Human U1 loci: genes for human U1 RNA have dramatically similar genomic environments. *Cell*, **29**, 257–264.

Maris,J.M., Jensen,J., Sulman,E.P., Beltinger,C.P., Allen,C., Biegel,J.A., Brodeur,G.M. and White,P.S. (1997) Human Kruppel-related 3 (*HKR3*): a candidate for the 1p36 neuroblastoma tumour suppressor gene? *Eur. J. Cancer*, **33**, 1991–1996.

Marzluff,W.F. (1992) Histone 3′ ends: essential and regulatory functions. *Gene Expression*, **2**, 93–97.

McCracken,S.N., Fong,N., Yankulov,K., Ballantyne,S., Pan,G.H., Greenblatt,J., Patterson,S.D., Wickens,M. and Bentley,D. (1997) The C-terminal domain of RNA polymerase II couples messenger RNA processing to transcription. *Nature*, **385**, 357–361.

Murphy,S. (1997) Differential *in vivo* activation of the class II and class III snRNA genes by the POU-specific domain of Oct-1. *Nucleic Acids Res.*, **25**, 2068–2076.

Murphy,S., Tripodi,M. and Melli,M. (1986) A sequence upstream from the coding region is required for transcription of the 7SK RNA genes. *Nucleic Acids Res.*, **14**, 9243–9260.

Nagai,K. and Mattaj,I.W. (1994) RNA–protein interactions in the splicing snRNPs. In Nagai,K. and Mattaj,I.W. (eds), *Frontiers in Molecular Biology*. IRL Press at Oxford University Press, Oxford, UK, pp. 150–177.

Neugebauer,K.M. and Roth,M.B. (1997) Transcription units as RNA processing units. *Genes Dev.*, **11**, 3279–3285.

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

Neuman de Vegvar,H.E. and Dahlberg,J.E. (1989) Initiation and termination of human U1 transcription requires the concerted action of multiple flanking sequence elements. *Nucleic Acids Res.*, **17**, 9305–9318.

Parry,H.D., Tebb,G. and Mattaj,I.W. (1989) The *Xenopus* U2 gene PSE is a single, compact, element required for transcription initiation and 3′ end formation. *Nucleic Acids Res.*, **17**, 3633–3644.

Pavelitz,T., Rusche,L., Matera,A.G., Scharf,J.M. and Weiner,A.M. (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the *RNU2* locus. *EMBO J.*, **14**, 169–177.

Sadowski,C.L., Henry,R.W., Lobo,S.M. and Hernandez,N. (1993) Targeting TBP to a non-TATA box *cis*-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev.*, **7**, 1535–1548.

Saito,T., Seki,N., Matsuda,Y., Kitahara,M., Murata,M., Kanda,N., Nomura,N., Yamamoto,T. and Hori,T.A. (1995) Identification of the human *ERK* gene as a putative receptor tyrosine kinase and its chromosomal localization to 1p36.1: a comparative mapping of human, mouse and rat chromosomes. *Genomics*, **26**, 382–384.

Seipelt,R.L., Zheng,B., Asuru,A. and Rymond,B.C. (1999) U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Res.*, **27**, 587–595.

Steinmetz,E.J. (1997) Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell*, **89**, 491–494.

Strobl,L.J. and Eick,D. (1992) Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-*myc in vivo*. *EMBO J.*, **11**, 3307–3314.

Terns,M.P., Dahlberg,J.E. and Lund,E. (1993) Multiple *cis*-acting signals for export of pre-U1 snRNA from the nucleus. *Genes Dev.*, **7**, 1898–1908.

Van Arsdell,S.W. and Weiner,A.M. (1984) Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell. Biol.*, **4**, 492–499.

Van der Drift,P., Chan,A., van Roy,N., Laureys,G., Westerveld,A., Speleman,F. and Versteeg,R. (1994) A multimegabase cluster of snRNA and tRNA genes on chromosome 1p36 harbours an adenovirus/SV40 hybrid virus integration site. *Hum. Mol. Genet.*, **3**, 2131–2136.

Van der Drift,P., Chan,A., Laurets,G., van Roy,N., Sickmann,G., den Dunnen,J., Westerveld,A., Speleman,F. and Versteeg,R. (1995) Balanced translocation in a neuroblastoma patient disrupts a cluster of small nuclear RNA U1 and tRNA genes in chromosomal band 1p36. *Genes Chromosomes Cancer*, **14**, 35–42.

Weber,J., Jelinek,W. and Darnell,J.E.,Jr (1977) The definition of a large viral transcription unit late in AD2 infection of HeLa cells: mapping of nascent RNA molecules labelled in isolated nuclei. *Cell*, **10**, 611–616.

Westin,G., Zabielski,J., Hammarstrom,K., Monstein,H.J., Bark,C. and Pettersson,U. (1984) Clustered genes for human U2 RNA. *Proc. Natl Acad. Sci., USA*, **81**, 3811–3815.

Whitelaw,E., Hogben,P., Hanscombe,O. and Proudfoot,N.J. (1989) Transcriptional promiscuity of the human α-globin gene. *Mol. Cell. Biol.*, **9**, 241–251.

Wieben,E.D., Nenninger,J.M. and Pederson,T. (1985) Ribonucleoprotein organisation of eukaryotic RNA. XXXII. U2 small nuclear RNA precursors and their accurate 3′ processing *in vitro* as ribonucleoprotein particles. *J. Mol. Biol.*, **183**, 69–78.

Yuo,C., Ares,M. and Weiner,A.M. (1985) Sequences required for 3′ end formation of human U2 small nuclear RNA. *Cell*, **42**, 193–202.

Received February 10, 1999; revised and accepted March 29, 1999