The EMBO Journal vol.4 no.7 pp.1827–1837, 1985

Formation of the 3′ end of U1 snRNA is directed by a conserved sequence located downstream of the coding region

Nouria Hernandez

Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

Communicated by W.Keller

U1 is a small non-polyadenylated nuclear RNA that is transcribed by RNA polymerase II and is known to play a role in mRNA splicing. The mature 3′ end of U1 snRNA is formed in at least two steps. The first step generates precursors of U1 RNA with a few extra nucleotides at the 3′ end; in the second step, these precursors are shortened to mature U1 RNA. Here, I have determined the sequences required for the first step. Human U1 genes with various deletions and substitutions near the 3′ end of the coding region were constructed and introduced into HeLa cells by DNA transfection. The structure of the RNA synthesized during transient expression of the exogenous U1 gene was analyzed by S1 mapping. The results show that a 13 nucleotide sequence located downstream from the U1 coding region and conserved among U1, U2 and U3 genes of different species is the only sequence required to direct the first step in the formation of the 3′ end of U1 snRNA.

Key words: transient expression/homologous system/mutated human U1 genes/U1 snRNA 3′ end formation

Introduction

In higher eukaryotic cells, the formation of the mature 3′ end of transcripts synthesized by RNA polymerase II is accomplished by at least two, and possibly more, distinct mechanisms. One mechanism generates the 3′ ends of polyadenylated mRNAs. All polyadenylated mRNAs studied so far are initially synthesized as precursors that extend beyond the polyadenylation site (Acheson, 1978; Ford and Hsu, 1978; Nevins and Darnell, 1978; Nevins *et al.*, 1980; Hofer and Darnell, 1981; Amara *et al.*, 1984; Frayne *et al.*, 1984; Hagenbüchle *et al.*, 1984; Mather *et al.*, 1984; Sheffery *et al.*, 1984). The 3′ end that will be the substrate for polyadenylation is thought to be generated by endonucleolytic cleavage of the precursor. This processing step requires a short signal, AAUAAA, located 10 to 30 nucleotides (nt) upstream of the polyadenylation site and, in some cases, sequences located downstream as well (Fitzgerald and Shenk, 1981; Higgs *et al.*, 1983; Montell *et al.*, 1983; Simonsen and Levinson, 1983; Gil and Proudfoot, 1984; McDevitt *et al.*, 1984; Wickens and Stephenson, 1984; McLauchlan *et al.*, 1985). The enzymes and factors that perform these reactions have not been identified. Because the small nuclear RNA (snRNA) U4 has short stretches of sequence complementary to the AAUAAA signal and to a tetranucleotide often found just downstream of the polyadenylation site, it has been proposed that the U4/U6 snRNP (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984) is somehow involved (Berget, 1984).

A second processing mechanism forms the 3′ end of some RNAs that do not contain an AAUAAA signal and that are not polyadenylated, namely those encoding the histones (Birchmeier *et al.*, 1984; Krieg and Melton, 1984; Price and Parker, 1984). This processing reaction requires two blocks of extremely well-conserved sequences that flank the mature 3′ end of the mRNA: a G,C-rich region with dyad symmetry which can fold the end of the mRNA into a stem-and-loop structure, and a purine-rich tract located about 15 nt further downstream (Birchmeier *et al.*, 1982, 1983; Georgiev and Birnstiel, 1985). Formation of the 3′ end of histone mRNA requires a 12S particle (Stunnenberg and Birnstiel, 1982), which belongs to the U-family of snRNPs (Lerner and Steitz, 1981) because it contains a U-rich RNA (called U7) and is precipitable by antibodies of the Sm specificity. The sequences of U7 cDNAs suggest that the 5′ end of U7 RNA can form base pairs with the two conserved blocks of sequence in the histone mRNA precursors (Strub *et al.*, 1984). Such recognition of a processing site through base-pairing is a strategy known to be employed by another snRNP, the U1 snRNP that uses the 5′ end of U1 snRNA to recognize 5′ splice sites during mRNA splicing (Mount *et al.*, 1983; Krämer *et al.*, 1984).

The U-snRNAs themselves, or at least U1 to U5 (Frederiksen *et al.*, 1978; Murphy *et al.*, 1982; Chandrasekharappa *et al.*, 1983) represent yet another class of non-polyadenylated RNAs synthesized by RNA polymerase II. The initiation of transcription of the U1 and U2 genes is directed by an unusual promoter (Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Ares *et al.*, 1985), but very little is known about how the 3′ ends of these RNAs are formed. At least two steps seem to be involved; precursors to U1 and U2 RNAs with several additional nucleotides at the 3′ end are generated, and these are then shortened to the size of the mature RNAs (Zieve and Penman, 1976; Eliceiri, 1980; Chandrasekharappa *et al.*, 1983; Madore *et al.*, 1984). Interestingly, the sequences surrounding the 3′ end of the genes encoding U1, U2, and U3 have features similar to those surrounding the 3′ end of histone genes. First, as in histone mRNAs, the 3′ ends of each of these U-RNAs can potentially fold into a hairpin structure, although the size and sequence of this structure is not conserved among the snRNAs. Second, a common block of sequence ($GTTTN_{0-3}AAAPuNNAGA$, see Table I) that bears some resemblance to the histone purine-rich box is located 9–19 nt downstream of the 3′ end of the RNAs (Mattaj and Zeller, 1983). Here, I show for U1 RNA that this 3′ element is the only sequence in the vicinity of the 3′ end of the gene that is absolutely required for the first step in the formation of the 3′ end of the RNA.

Results

Transient expression of a marked human U1 gene in HeLa cells

A human U1 gene, flanked by 425 nt of 5′ sequence and 92 nt of 3′ sequence, was modified to encode a marked U1 snRNA that could be distinguished from the endogenous U1 snRNA in HeLa cells. An *Eco*RI linker was inserted into the *Bcl*I site 27 nt downstream from the 5′ end of the gene, and a *Taq*I site located 50 nt upstream of the 3′ end was converted to a *Sal*I site (for

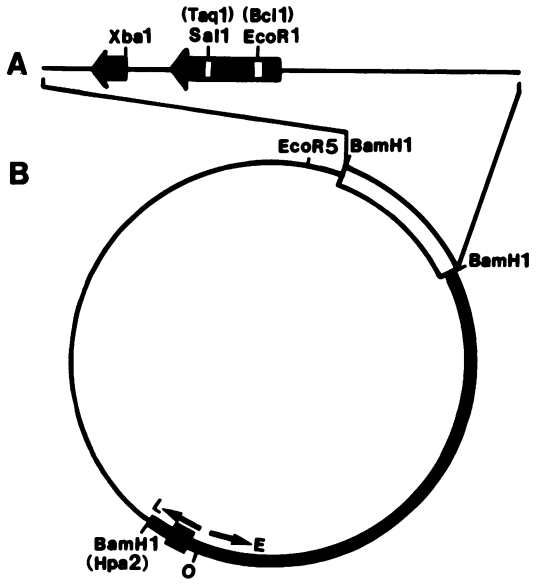

Fig. 1. Structure of pRec2, the parental plasmid. **A**: structure of the insert. Black boxes represent U1 coding sequences. The arrows indicate the direction of transcription and mark the 3′ end of the gene. The white areas in the coding sequence represent small insertions: 6 nt inserted in the *Taq*I site create a *Sal*I site, and 14 nt inserted in a *Bcl*I site create an *Eco*RI site. The thin line represents U1 flanking sequences. **B**: structure of the expression vector. The thin line corresponds to pBR322 DNA with two deletions, from nt 1426 to nt 2525, and from nt 3102 to nt 3211. The white box represents the inserted human DNA. The black box represents SV40 sequences, from the *Bam*HI site at nt 2533 to the destroyed HpaII site at nt 346. The location of the enhancer is indicated by a thickened black area.

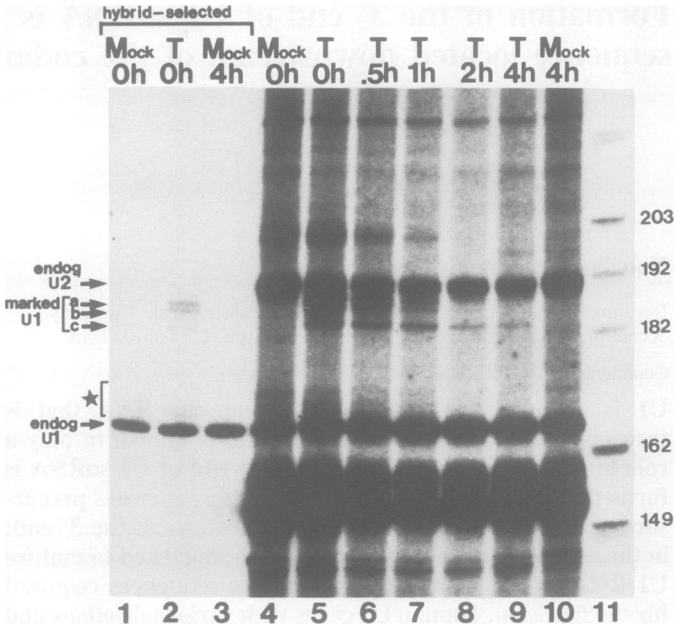

Fig. 2. Expression of the marked U1 gene in HeLa cells. HeLa cells were labeled with [^{32}P]phosphate for 12 h. The medium was then replaced by medium containing cold phosphate and 0.1 μg/ml of actinomycin D, a concentration that completely blocks synthesis of U1 RNA in HeLa cells (Zieve *et al.*, 1977). RNA was extracted immediately (**lanes 1, 2, 4, 5**) or after 0.5 h (**lane 6**), 1 h (**lane 7**), 2 h (**lane 8**), 4 h (**lanes 3, 9, 10**). **Lanes 2** and **5–9**: RNA from transfected cells. **Lanes 1, 3, 4** and **10**: RNA from mock-transfected cells. Bands corresponding to endogenous U1 and U2 RNA, and to marked U1 RNAs, are indicated. Bands corresponding to U1 RNAs slightly longer than mature endogenous U1 are denoted by a star. **Lane 11**: DNA size markers.

details, see Materials and methods). The length of the encoded RNA was thereby increased from 164 to 184 nt. The marked gene was introduced into a eucaryotic expression vector (pSVE) that contained the entire early region of SV40, as well as the transcriptional enhancer. The last 50 nt of the U1 gene together with the 90 nt of 3′ flanking sequence were then duplicated to produce tandem sites at which mature 3′ ends could be formed (see Figure 1). The first site (site I) was the target of the mutations described below; the second (site II) ensured that RNAs with discrete 3′ ends would be formed in case a mutation inactivated the first site.

This construct (called pRec2) was introduced into HeLa cells by transfection of calcium phosphate precipitated DNA, and RNA was collected 44–50 h later. The efficiency of transfection, monitored by indirect immunofluorescence with a monoclonal antibody directed against SV40 T antigen, was between 30 and 70% (data not shown). To determine first whether U1 RNA was synthesized from the exogenous U1 gene, transfected and mock-transfected HeLa cells were labeled with [^{32}P]phosphate for 12 h beginning 32 h after transfection. The cells were then washed and transferred to medium containing unlabeled phosphate and 0.1 μg/ml of actinomycin D, a concentration that completely blocks synthesis of U1 RNA in HeLa cells (Zieve *et al.*, 1977). RNA was extracted immediately or after a chase period of 0.5 h, 1 h, 2 h or 4 h, and then electrophoresed on a 4% polyacrylamide-urea gel. As shown in Figure 2, transfected cells synthesize three RNA species that are absent from mock-transfected cells (compare lanes 4 and 5). These RNAs (labeled a, b, and c in Figure 2) can be hybrid-selected by a single stranded M13 recombinant DNA containing the anti-sense strand of the U1 gene (lane 2), showing that they are U1 transcripts. Together, they are produced at a level ~10% that of the endogenous U1. The shortest of these RNAs (species c) is about 184 nt long, as expected for a correctly initiated and terminated transcript of the marked U1 gene. The longer RNAs (species a and b) contain from 3 to 5 additional nucleotides. All three RNAs are much less stable than endogenous U1 RNA. Species c has a half-life of 1–2 h (lanes 5 to 9), while species a and b disappear even faster, consistent with their being precursors of the shortest species. The RNAs slightly longer than endogenous U1 (indicated by a star in Figure 2) follow similar kinetics and may be precursors of U1 RNA.

Because the RNAs derived from the marked U1 gene have a correct and unique 5′ end (see below), the length heterogeneity must result from a variation at the 3′ end. To confirm this, S1 mapping was performed separately on RNA species a+b and species c, which had been excised from a gel like that shown in Figure 2. The probe was a single-stranded fragment 3′ end-labeled at the *Eco*RI site within the U1 gene, and extending through the tandem termination sites and into pBR322 sequences. When this probe was annealed to species c, only one protected DNA fragment was obtained (Figure 3, lane 5). By comparison with a chemical sequencing ladder derived from the probe, this fragment was one to two nt shorter than expected if protected by an RNA having the same 3′ end as mature endogenous RNA. This was due to S1 nibbling at the end of the hybrid, since S1 mapping of endogenous U1 under the same conditions also gave a protected DNA fragment slightly shorter than expected (not shown). Moreover, when the S1 concentration was reduced, additional bands were detected that were longer by one or two nucleotides (see Figure 5 for an example). When the probe was annealed to RNA species a+b, two protected fragments were obtained (lane 4). The slower-migrating fragment was 4 nt longer than the fragment protected by species c, showing that a and b

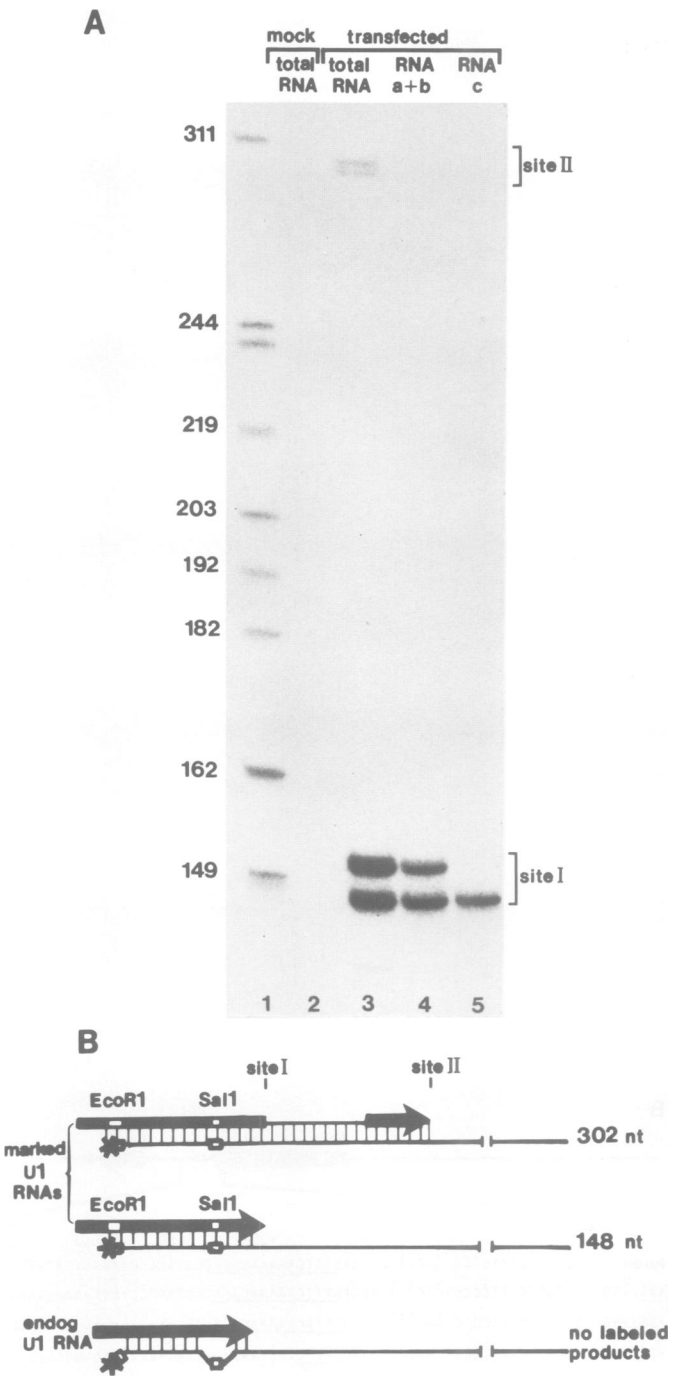

Fig. 3. S1 analysis of RNAs synthesized from the marked U1 gene. **A**: a single stranded DNA probe 3′ end-labeled at the *Eco*RI site in the coding region of the marked U1 gene was annealed to RNA extracted for mock-transfected cells (**lane 2**) or transfected cells (**lane 3**), and to RNAs a+b (**lane 4**) or RNA c (**lane 5**) eluted from a gel similar to that displayed in Figure 2. The hybrids were treated with nuclease S1 and the protected DNA fragments were fractionated on a 6% sequencing gel. **Lane 1**: DNA size markers. **B**: structure of the RNA-DNA hybrids. The top lines of the three hybrids represent marked U1 RNA ending at site II, marked U1 RNA ending at site I, and endogenous U1 RNA. The bottom line represents the DNA probe, with its labeled 3′ end indicated by a star and the sequences corresponding to the *Eco*RI and *Sal*I sites inserted in the coding sequence of U1 indicated by small boxes. The expected lengths of probe protected by the different RNAs are given on the right.

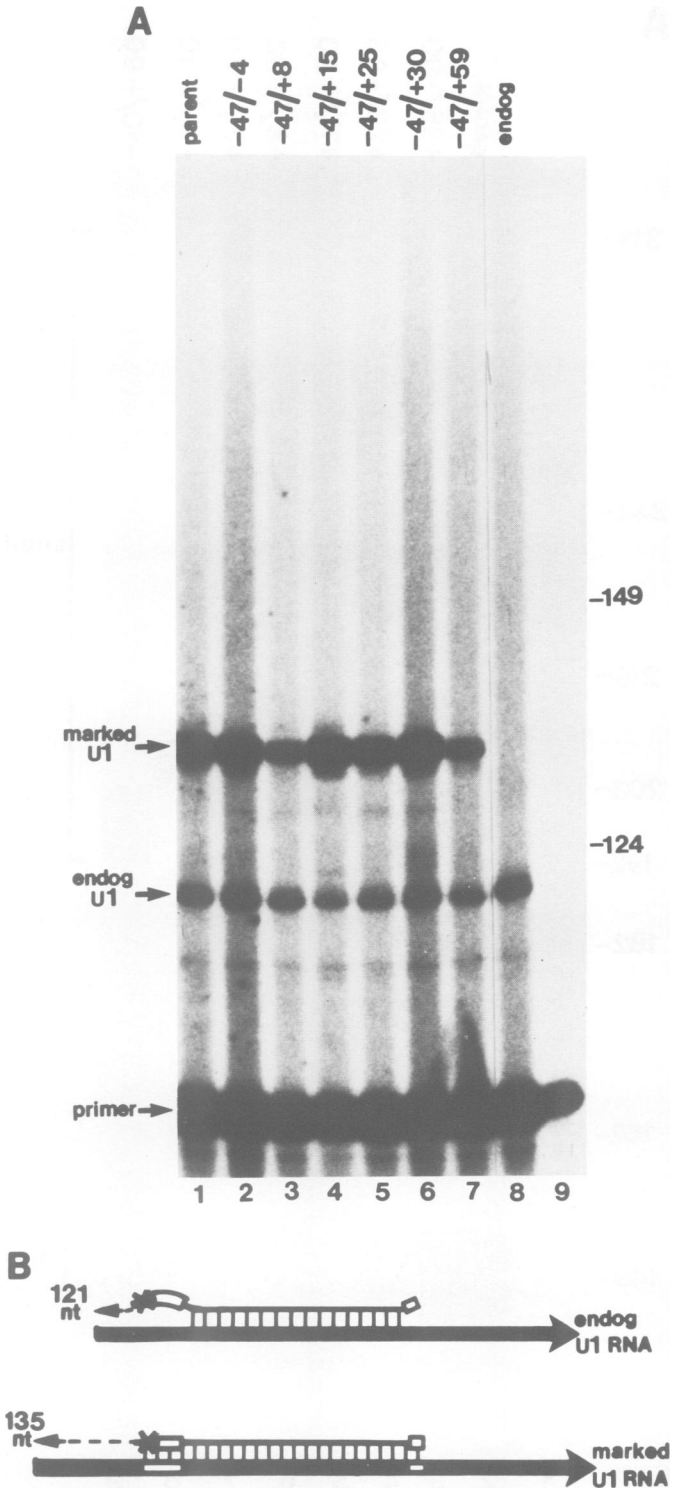

Fig. 4. Primer extension analysis of RNAs synthesized from the marked U1 gene. A labeled primer extending from the *Eco*RI to the *Sal*I site in the U1 gene was hybridized to RNA extracted from cells transfected with pRec2, the construct containing the full-length marked U1 gene (**lane 1**), or with constructs containing various deleted marked U1 genes (**lanes 2–7**) whose structures are shown in Figure 6B. In **lanes 8** and **9,** the primer was hybridized to RNA extracted from mock-transfected cells and to tRNA, respectively. The primer was extended with AMV reverse transcriptase and unlabeled deoxynucleotide triphosphates, and the cDNA products were fractionated on a 6% sequencing gel. cDNAs corresponding to endogenous and marked U1 RNAs are indicated, as well as the position of DNA size markers. **B**: structure of the RNA-DNA primer hybrids. The labeled 3′ end of the primer is indicated by a star. The white boxes represent the *Eco*RI and *Sal*I sites inserted in the coding region of the U1 gene. The expected length of the cDNA products is indicated on the left.

are longer at their 3′ ends. Surprisingly, the faster-migrating fragment had the same length as the one protected by species c. Again this is presumably due to S1 nibbling at the end of the longer

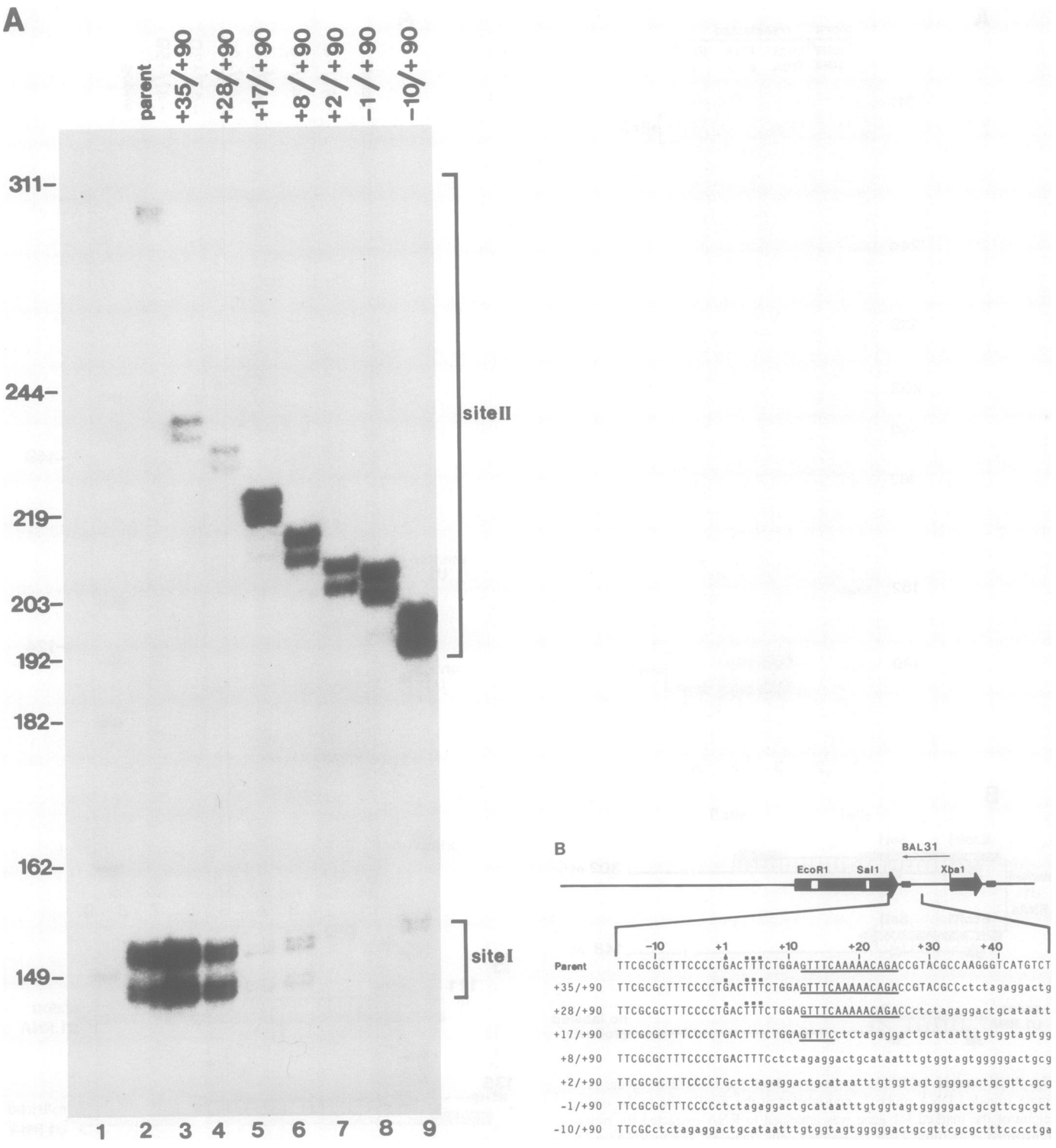

Fig. 5. Analysis of expression and structure of mutant U1 genes from the *Xba* series. **(A)** S1 analysis of RNA extracted from cells transfected with pRec2 (**lane 2**) or with various deleted marked U1 genes (**lanes 3–9**) whose structure is displayed in part B. The DNA probes were single-stranded restriction fragments similar to the one used in Figure 3, isolated from each different construct and labeled in parallel, so as to have comparable specific activities. Bands corresponding to RNAs ending at site I and site II are indicated. **Lane 1**: DNA size markers. **(B)** Structure of the mutant genes. On the top, the start site (*Xba*I site) and the direction of the BAL31 deletions are indicated. The small black boxes represent the conserved downstream element of the U1 gene. On the bottom, the sequence around the endpoints of the deletions is displayed. The underlined nucleotides constitute the downstream conserved element. Nucleotide +1 is the last nucleotide of mature U1 RNA. The location of 3′ ends is indicated by small squares above the sequence. The positions are approximate, since S1 nibbling at the ends of the DNA-RNA hybrids prevents an exact mapping of the 3′ ends.

hybrids. Total RNA from transfected cells protected mainly the same two fragments and, to a much lesser extent, two longer fragments that correspond to RNAs ending at site II (lane 3). As expected, RNA extracted from mock-transfected cells gave no signal, since the labeled end of the probe from the marked gene cannot be protected by endogenous U1 RNA (lane 2 and

see Figure 3B). Thus, the marked U1 gene directs the transcription of three major RNA species that differ at their 3′ ends and that protect two characteristic DNA fragments from digestion by S1.

28 nt of 3′ flanking sequence are sufficient for formation of the 3′ end of U1 RNA

In order to equalize the amount of RNA from different transfections, primer extension was used to monitor the amount of correctly initiated U1 RNA in each preparation. The primer extended from the *Eco*RI site to the *Sal*I site of the marked U1 gene (Figure 4B), and was hybridized to a fixed percentage of RNA from each transfection. Two major cDNAs were obtained (Figure 4, lanes 1−8). The longer has the size expected of a cDNA of the marked U1 RNA initiated at the natural cap site (135 nt). The shorter (121 nt) corresponds to endogenous U1, and is the only U1 RNA detected in non-transfected cells (lane 8). The bands from the marked U1 RNAs were excised and their radioactivity determined. Samples were normalized so that equal amounts of correctly initiated RNA were then used for S1 analysis. (Note that because 7 nt at the 3′ end of the primer are not complementary to endogenous U1 RNA, the primer is not efficiently extended on this template. For this reason, the relative intensities of the signals derived from endogenous U1 RNA and marked U1 RNA do not reflect the ratio of their cellular concentration (see Figure 4B)).

The sequences required to form the 3′ end of U1 RNA were determined by introducing systematic mutations in the 3′ end region of the gene. The mutated genes were then transfected into HeLa cells, and the synthesized RNA was analyzed by S1 mapping. I first tested whether the formation of a correct 3′ end required sequences in the region downstream of the U1 coding sequence. A series of BAL31 deletions (series Xba) was constructed that began at the *Xba*I site, 90 nt downstream from the 3′ end of mature U1, and progressively removed sequences in a 3′ to 5′ direction (see Figure 5B). The *Xba*I restriction site and the auxiliary site of 3′ end formation (site II) were thus moved closer and closer to the 3′ end of the U1 gene. The deletions were designated by their endpoints, with the last nucleotide of mature marked U1 RNA numbered +1. Thus, deletion +35/+90 removes the sequence from nt 35 to nt 90 downstream of the 3′ end (see Figure 5B). RNA from each construct was subjected to S1 analysis with colinear 3′ end-labeled probes that were similar to the one used in Figure 1. The signal given by RNA extracted from cells transfected with the parental construct is displayed in lane 1. Because digestion was performed with half the amount of S1 used in the experiment shown in Figure 3, two pairs of bands, rather than two single bands, were obtained. Again, only a small fraction of the transcripts ended at site II. Deletions that removed the 3′ flanking sequences to within 9 or 2 nt of the conserved 3′ element had no effect on the formation of 3′ ends (lanes 3 and 4). By contrast, deletion +17/+90, which removed part of the conserved 3′ element, or deletions extending closer to the 3′ end, reduced the efficiency of 3′ end formation at site I by at least 200 fold. Concomitantly, a large fraction of the RNA terminated at site II (lanes 5−9). These results demonstrate that 28 nt of 3′ flanking sequences are sufficient to promote correct and efficient formation of U1 3′ ends.

The 3′ terminal stem-and-loop of U1 is not required for 3′ end formation

The second series of BAL31 deletions, series *Sal*1, was constructed to determine whether the 3′ terminal stem-and-loop and the nucleotides surrounding the 3′ end of the U1 gene were required for 3′ end formation. The deletions extended from position −47, the last nt of the *Sal*I site in the marked U1 coding sequence, to positions closer and closer to the 3′ end; the largest deletions removed the site of 3′ end formation itself, and portions of the 3′ flanking sequence (Figure 6B).

An S1 analysis of the RNAs synthesized by these different mutants is displayed in Figure 6A. When internal sequences from nt −47 to nt −14, −10, −7 or −6 were deleted from the marked U1 gene, the two protected DNA fragments diagnostic of RNAs with correct 3′ ends were observed, only shortened by the extent of the internal deletions (lanes 1−4). These deletions removed extensive portions of the region of dyad symmetry at the 3′ end; in deletion −47/−14, no stable hairpin structure could be formed within the 20 nt preceding the 3′ end. To exclude the possibility that sequences that are located upstream from the *Sal*I site and that were moved closer to the 3′ end by the BAl31 deletions were functionally replacing the deleted sequences, genes with substitutions were constructed (R series). Fragments from members of the *Xba* series that extended from the 5′ end of the gene to the *Xba*I site were ligated to fragments isolated from members of the *Sal* series that extended from the *Sal*I site to the end of the 3′ flanking sequences. In these mutants, internal U1 sequences near the 3′ end of the coding region were replaced by a 12 nt sequence corresponding to the filled *Xba*I site ligated to the filled *Sal*I site. In mutants R−30/−10, R−30/−7 and R−30/−6, U1 sequences between nt −30 and −10, −7 or −6 were deleted and replaced by the 12 nt sequence (Figure 7B). These mutants also directed the synthesis of RNAs with correct 3′ ends, (Figure 7A, lanes 2, 3, 4).

Sequences around nt +1 direct the accuracy of the reaction

The mutant −47/−4 in Figure 6B and R−10/−4 in Figure 7B retained the sequence encoding the last 4 nt of mature U1 RNA. In addition to RNAs with wild-type 3′ ends, these genes produced low amounts of RNAs with new 3′ ends, as testified by the appearance of additional shorter DNA fragments protected from S1 digestion (Figure 6A, lane 5; Figure 7A, lane 5). These new 3′ ends were located in a 20 nt segment just upstream of site I. Furthermore, more extensive deletions that removed site I completely (−47/+8), or site I together with the first 4 nt of the conserved downstream element (−47/+15) resulted in synthesis of RNAs with heterogeneous 3′ ends that mapped in a region 8 to 16 nt upstream of the conserved element (Figure 6A, lanes 6 and 7). These results show that the DNA sequence that encodes the last 5 or 6 nt of mature U1 RNA is required for accurate formation of 3′ ends. By contrast, neither this sequence nor the 14 nt immediately downstream of nt +1 contribute to the efficiency of the reaction.

Sequences downstream of nt +15 are required to form 3′ ends

Deletions −47/+25, −47/+30 and −47/+59 removed both site I and the conserved downstream element together, with the result that RNAs ended at site II (Figure 6A, lanes 8, 9, 10).

The *Sal* deletion series indicates that sequences upstream of nt +15 are dispensable, while the *Xba* series shows that sequences downstream of nt +28 are also dispensable. Nt +15 and nt +28 thus mark the boundaries of an element that is essential for generating the 3′ end of U1 RNA. This segment coincides nearly exactly with the conserved downstream element previously identified by phylogenetic comparisons (Mattaj and Zeller, 1983). To confirm the importance of this element, three more mutants belonging to the R series were constructed (Figure 8B). Mutant R+8/+25 replaces the entire conserved downstream element by the 12 nt sequence from the fused *Xba*I

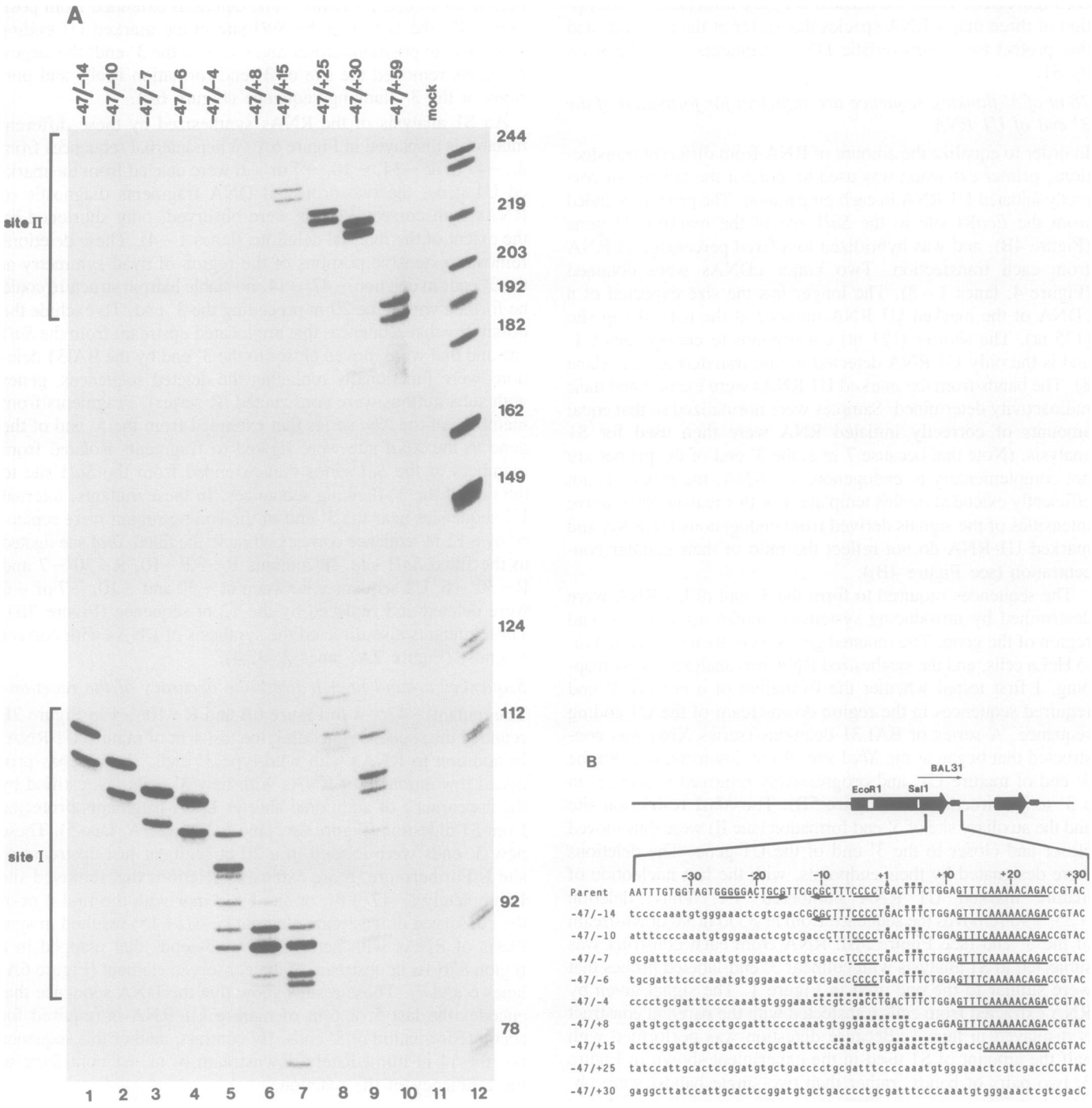

Fig. 6. Analysis of expression and structure of mutant U1 genes from the *Sal* series. (**A**) S1 analysis of RNAs extracted from cells transfected with various deleted marked U1 genes (**lanes 1–10**) whose structure is displayed in part B, and of RNA extracted for mock-transfected cells (**lane 11**). **Lane 12**: DNA size markers. Bands corresponding to RNAs ending at site I and site II are indicated. (**B**) Structure of mutant genes. On the top, the start site (*Sal*I site) and the direction of the BAL31 deletions are indicated. On the bottom, the sequence around the endpoints of the deletions is displayed. The terminal dyad symmetry of the U1 gene is underlined by arrows. For other symbols, see Figure 5B.

and *Sal*I sites. Mutant R+17/+25 replaces only the last 10 nt of the conserved element. As shown in Figure 8A, both mutations resulted in a 200-fold reduction of U1 RNAs ending at site I and a corresponding increase in RNAs ending at site II (lanes 3 and 4). Thus, in the absence of the conserved downstream element, 3′ end formation was virtually abolished.

Mutant R+8/+15 has lost the first 4 nt of the conserved downstream element and the 4 nt upstream of the element. Nevertheless, RNA transcribed from this gene was efficiently and accurately terminated at site I (lane 2). The first 4 nt of the conserved downstream element, with the possible exception of the first G which is restored by the substituting sequence, are therefore dispensable. This was an expected result, since mutant Sal−47/+15 had earlier proved capable of directing efficient formation of 3′ ends, although the 3′ ends were heterogeneous. Indeed, in that case the site of 3′ end formation itself had been deleted (Figure 6A and 6B).

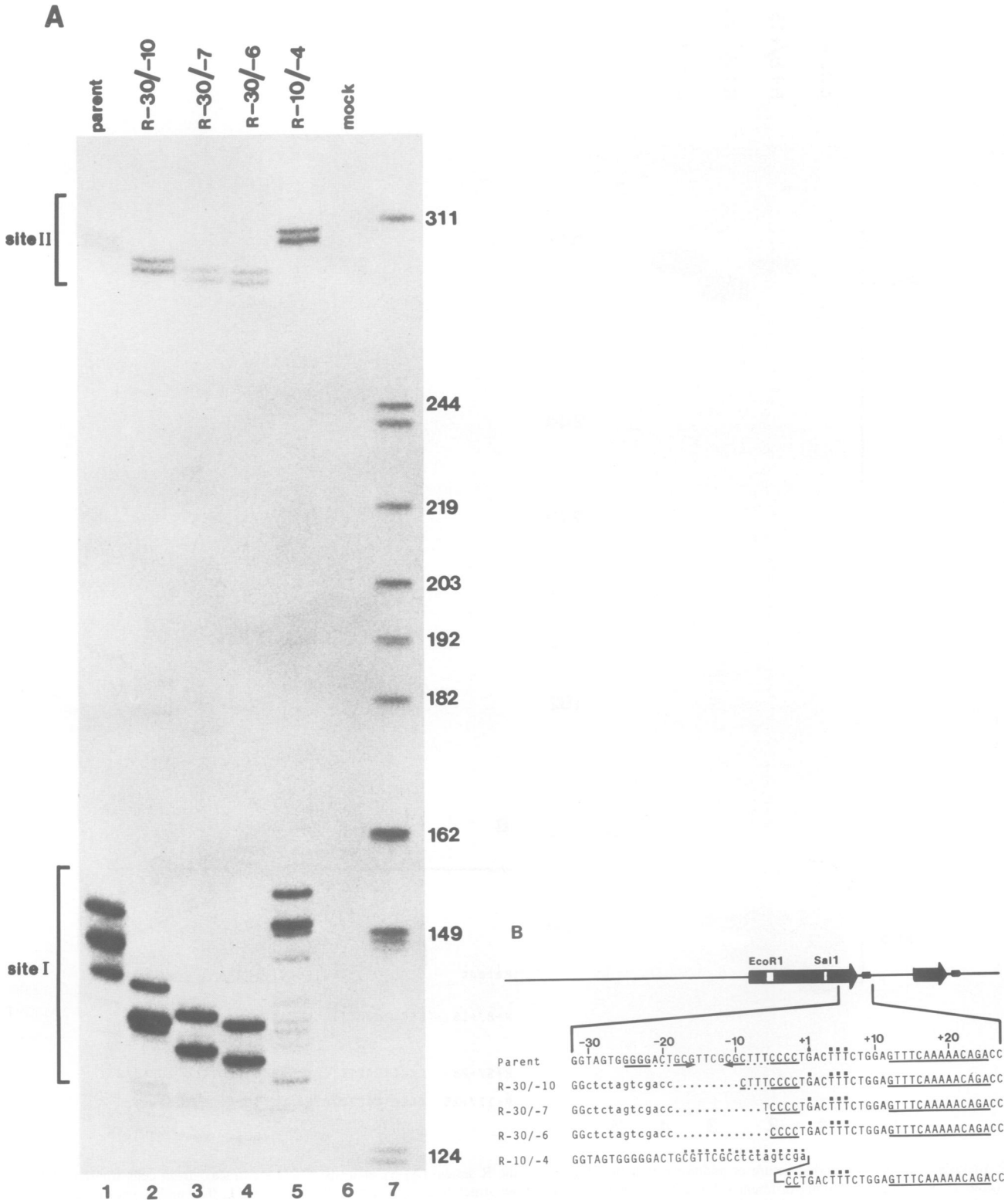

Fig. 7. Analysis of expression and structure of mutant U1 genes from the R series. (**A**) S1 analysis of RNA extracted from cells transfected with pRec2 (**lane 1**) or with genes having different substitutions whose structure is shown in part B. In **lane 6**, S1 analysis of RNA from mock-transfected cells. **Lane 7**: DNA size markers. (**B**) Structure of the substituted genes. The 12 nucleotides that replace part of the deleted sequence are represented by lower-case letters. The deleted nucleotides that are not substituted are marked as dots. In R−10/−4, fewer nucleotides were deleted than substituted. Symbols are the same as in Figure 5B.

Discussion

Expression of an exogenous human U1 gene in HeLa cells

This study explores the sequences required to form the 3′ end of U1 RNA. The experiments used HeLa cells as a homologous system for transient expression of a human U1 gene linked to the transcriptional enhancer of SV40. The exogenous U1 gene

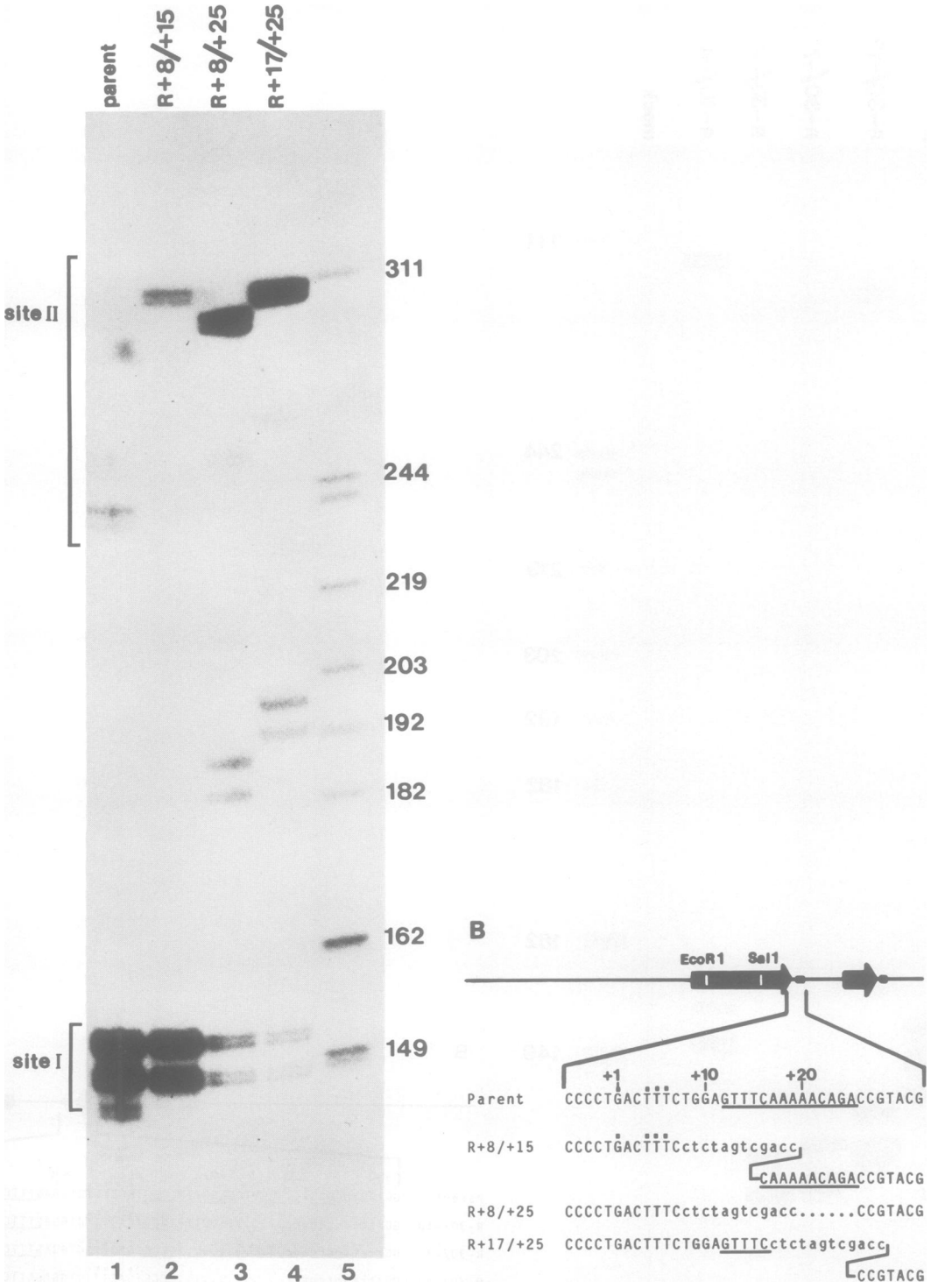

Fig. 8. Analysis of expression and structure of additional mutant U1 genes of the R series. (**A**) S1 analysis of RNA extracted from cells transfected with pRec2 (**lane 1**) or with genes having different substitutions (**lanes 2, 3, 4**) whose structure is shown in part B. (In **lane 1**, the bands corresponding to RNAs ending at site II are shorter because the S1 probe used in this case was not isolated from pRec2, but from the mutant +35/+90.) **Lane 5**: DNA size markers. (**B**) Structure of the substituted genes. Symbols are as in Figure 7B.

was marked by the insertion of two new restriction sites, so that its transcripts could be distinguished from the endogenous U1 species.

The marked U1 gene produced three major RNAs. The shortest is the fully mature form; the two others were 3–5 nt longer at their 3′ ends. U1 RNAs slightly longer than the endogenous U1 RNA were detected as well (Figure 1), which were short-lived, like their marked counterparts. In both cases, the different stabilities of these RNAs suggest that the longer are precursors to the shortest. This precursor-product relationship is supported by previous observations of cytoplasmic precursors to U1 and U2 RNAs (Eliceiri, 1974; Zieve and Penman, 1976), although no consensus about the length of these precursors has been reached. Eliceiri and Sayavedra (1976) and Eliceiri (1980) reported precursors with about three extra nt while Madore *et al.* (1984) observed precursors with one to 15 extra nt. Perhaps these

Table I. Conserved element downstream of snRNA genes and some histone genes

Human U1							
(HSD1, HSD6)	ACTTTCTGGA	GTTT	C- -	AAAA	AC	AGA	CTG
(HSD2,4,5,7)	ACTTTCTGGA	GTTT	C- -	AAAA	AC	AGA	CCT
(HSD3)	ACTTTCTGGA	GTTT	C- -	AAAA	GT	AGA	CTG
Human U2	CCTCCGGGGATACAACGT	GTTT	CCT	AAAA	GT	AGA	GGG
Mouse U1	ATTTTTGTG	GTGC	TA-	AAAG	TT	AGA	TGC
Mouse U2	CCTCTGGGGAATAAGAATG	GTCA	G- -	AAAA	AT	AGA	CTT
Rat U1	GCATTTCTG	GTAT	GAG	AAAG	TA	AGA	GTT
Rat U2	CCTCCGGGGAATAATAACT	GTTT	- - -	AAAG	TT	AGA	AGT
Rat U3B.4	CCCTCATACAATTATG	GTTT	TT-	AAAG	GA	AGA	TCT
Rat U3B.7	CTTCGGTCTAACCTTTATG	GTAT	TA-	AAAG	GA	AGA	GGC
Rat U3D	TTCTTTTTATTCCTT	GTTT	A- -	AAGG	TT	AGA	ACT
Chicken U1	ATTTGCGCG	GTTC	- - -	AAAG	AC	AGA	ACG
Xenopus U1	TTTTGGTTT	GTTT	- - -	AAAG	AT	AGA	GAT
Xenopus U2	CTCTTACTCA	GTTT	GA-	AAAA	GC	AGA	AAA
Drosophila U1	CATCTGCAATAA	ATTT	T- -	AAAA	GT	ATC	ATT
Drosophila U2	GAATAATAAATATTTAACT	ATTG	T- -	AAAA	TC	AGA	AGG
Sea urchin U1	CAATTCAACTA	ATGC	- - -	AAAG	AA	AGA	AAA
Consensus		GTTT	N_{0-3}	AAA^{G}_{A}	NN	AGA	
Sea urchin H3	ACCACA	CCCC	- - -	CAAG	AA	AGA	TTC
Human H2A	ACCTACGA	CTTC	CAT	TAAA	TG	AGC	TGT
Human H2B	ACCCAT	GTTC	TC-	AAAG	AA	AGA	GCT
Human H3	ACCCACT	TATT	CC-	AACG	AA	AGT	AGC
Human H4	CCTACTTCCTCA	GCTG	- - -	AAGG	TG	GTA	ACA

The snRNA gene sequences begin with the first nucleotide following the 3′ end of the mature RNA; the histone gene sequences begin with the first nucleotide following the 3′ dyad symmetry. Sequences of the different genes are from: human histones: Zhong *et al.*, 1983; human U1: Manser and Gesteland, 1982; human U2: Ares *et al.*, 1985; mouse U1: Marzluff *et al.*, 1983; mouse U2: Nojima and Kornberg, 1983; rat U1: Watanabe-Nagasu *et al.*, 1983; rat U2: Tani *et al.*, 1983; rat U3: Stroke and Weiner, 1985; chicken U1: Roop *et al.*, 1981; *Xenopus* U1: Zeller *et al.*, 1984; *Xenopus* U2: Mattaj and Zeller, 1983; *Drosophila* U1 and U2: Alonso *et al.*, 1984a, 1984b; sea urchin U1; Brown *et al.*, 1985; sea urchin H3: Hentschel and Birnstiel, 1981.

discrepancies reflect a mechanism of maturation which involves a 3′ to 5′ exonuclease, and which therefore produces many intermediates. Although a formal precursor-product relationship has not been established for the marked U1 RNAs reported here, for the purposes of this discussion I will use the term 'precursor' to refer to the longer RNAs.

Endogenous U1 RNA precursors are present at very low levels compared with mature U1 RNA, but in transfected cells the marked precursors make up at least half of the RNA synthesized from the exogenous gene. This is due in part to the surprisingly low stability of the mature RNA from the marked gene. The half-life of the marked U1 is only about 90 min, whereas endogenous U1 RNA appears to be metabolically stable (Weinberg and Penman, 1969). The basis for this difference is unclear. Although the mature, marked RNA can be precipitated by anti-Sm and anti-(U1) RNP antibodies (not shown), it might nevertheless be improperly or slowly assembled into a snRNP as a consequence of the modifications in its sequence. This in turn might slow its transport back into the nucleus (Madore *et al.*, 1984), thereby increasing the time it spends in the cytoplasm and, perhaps, the extent of its degradation. While the cellular location of the different marked U1 RNAs was not carefully examined, both mature U1 and U1 precursors were indeed found almost exclusively in the cytoplasmic fraction after aqueous fractionation (data not shown). A secondary contribution to the relative abundance of marked U1 precursors may come from slow maturation of precursors synthesized from the exogenous template.

Sequences required to form the 3′ ends of the precursors of U1 RNA

Changes in the sequences near the 3′ end of the U1 gene affected equally the amounts of the precursors and the mature, marked U1 RNAs. I assume, therefore, that these changes disrupted a reaction that forms the precursors. The sequence requirements for the second reaction, in which the precursors are fully matured, remain unknown.

Which sequences are necessary to form the 3′ ends of the precursors? Deletion analysis showed that the 3′ terminal hairpin just upstream of the mature 3′ end, as well as sequences more than 28 nt downstream of the 3′ end, are dispensable. So too are the few nucleotides where the 3′ ends normally map. When these nucleotides and the sequences immediately surrounding them were changed, however, RNAs with 3′ ends scattered over about 15 nt were observed, some of them perhaps new precursors, other improperly matured precursors. The region very near the mature 3′ end appears, therefore, to affect the accuracy of

the reaction. The only sequences that are absolutely required to form the 3′ ends of the precursors lie within the 16 nt sequence GTTTCAAAAACAGACC, which is found 10 nt downstream of the mature 3′ end. Of this 3′ element, the second, third and fourth nucleotide can be substituted without apparent deleterious effect. All or part of the remaining 13 nt therefore seem to constitute a functional unit that directs formation of 3′ ends in a short region upstream of itself. A similar 3′ element is involved in the generation of the 3′ end of human U2 RNA in *Xenopus* oocytes (Yuo *et al.*, 1985). It remains possible that sequences between the 5′ end of the marked U1 gene and the internal *Sal*I site, 130 nt farther downstream, are also necessary to form RNA 3′ ends. This seems unlikely, however, since these sequences are not conserved in different snRNA genes. Sequences upstream of the 5′ end of the gene are dispensable, because the U1 promoter can be replaced by the promoter of the herpes virus thymidine kinase gene without substantially perturbing the formation of 3′ ends (data not shown). Clearly, inserting the 3′ element in a completely new environment will establish whether it can function autonomously.

What kind of reaction does the 3′ element direct? One possibility is that it terminates transcription, either as a DNA or an RNA signal, although it has none of the features (dyad symmetry, run of Ts) associated with known terminators. Another possibility is that it is an RNA signal that directs a processing reaction, such as cleavage of a long primary transcript. The 3′ ends of polyadenylated mRNAs and of sea urchin histone mRNAs are formed in this way (see Birnstiel *et al.*, 1985). The 3′ element identified here, which is well-conserved among snRNA genes (see Table I), bears no resemblance to the AAUAAA sequence required to form the 3′ end of polyadenylated mRNAs, but a comparison with one of the critical 3′ elements involved in the formation of the 3′ end of sea urchin histone mRNAs, the CAAGAAAGA box, leads to a more ambiguous conclusion. Indeed, the sea urchin histone box can be aligned to match part of the U1 3′ element (Table I). However, although the sequences required to form the 3′ end of human histone mRNAs have not been determined, no such strongly conserved sequence as the sea urchin CAAGAAAGA box can be found downstream of the genes that have been isolated. But an almost exact match to the conserved 3′ element of snRNA genes can be found downstream of the human H2B gene. In other cases (human H2A, H3 and H4 genes) the match is less good (Table I). These sequence similarities are intriguing, and may indicate that the 3′ ends of snRNAs are formed by an RNA processing reaction that uses factors also involved in the processing that forms histone mRNAs. However, processing of histone RNAs must be different in some way, since, at least in the cases studied in detail (sea urchin H2A and H3 genes), the CAAGAAAGA box alone is not sufficient to form the 3′ ends of the mRNAs. Instead, the CAAGAAAGA box functions coordinately with the stem-and-loop structure found at the 3′ end of the mature mRNA. The CAAGAAAGA box is thus only part of a bipartite element, whose linear spacing is rigidly fixed (Georgiev and Birnstiel, 1985). This difference is supported by sequence comparisons; the sequence of the terminal stem-and-loop is very well-conserved among different histone mRNAs species (Hentschel and Birnstiel, 1981; Birchmeier *et al.*, 1983), whereas the 3′ stem-and-loops of snRNAs such as U1, U2 and U3 are unrelated to one another in size and sequence. The two parts of the histone bipartite element can potentially base-pair with U7 snRNA, a constituent of a snRNP involved in the processing reaction (Stunnenberg and Birnstiel, 1982; Strub *et al.*, 1984). None of the factors involved in the formation of the 3′ ends of snRNAs have been identified. The development of an *in vitro* system that can generate the 3′ ends of snRNAs will be useful to determine whether 3′ end formation is truly an RNA processing reaction, and whether it uses the U7 snRNP and/or other factors.

Materials and methods

Plasmid constructions

From a genomic clone (HSD4) originally isolated by Manser and Gesteland (1982), a 681 nt *Hae*III fragment containing a human U1 gene was fitted with *Bam*HI linkers and subcloned into the *Bam*HI site of pBR322, and unknown sequences were determined (M.Mangin, unpublished). A 10 nt *Eco*RI linker was inserted into a *Bcl*I site, 27 nt downstream of the 5′ end of the U1 gene, that had been filled in with Klenow polymerase. A *Taq*I site, 50 nt upstream of the 3′ end of the gene, was changed into a *Sal*I site as follows; the *Bam*HI fragment containing the U1 gene was cleaved with *Taq*I. Each of the two resulting fragments were mixed with the large fragment obtained by cleavage of pBR322 with *Bam*HI and *Sal*I, and ligase was added to join the sticky ends (pBR322 *Bam*HI site to the U1 *Bam*HI site). After 1 h, Klenow polymerase and deoxynucleotide triphosphates were added, and the incubation was continued overnight to allow ligation of the filled-in *Taq*I sites to the filled-in *Sal*I sites, generating a *Sal*I site immediately adjacent to the former *Taq*I site in the U1 gene. The U1 gene was then reconstructed by recloning the *Bam*HI-*Sal*I fragments containing the 5′ and 3′ portions of the gene into the *Bam*HI site of a pBR322 vector in which the *Sal*I site had been destroyed by filling in. Lastly, an *Xba*I linker was inserted into the *Bam*HI site downstream of the U1 gene. This construct, pNS11Eco$^+$SalXba, was the substrate for BAL31 deletions.

pRec2 was constructed as follows. The *Bam*HI fragment containing the marked U1 gene was inserted into a *Bam*HI site of pSVE, a plasmid containing the early region of SV40 including the transcriptional enhancer (obtained from David Solnick) (see Figure 1). To duplicate the 3′ end of the U1 gene, a *Sal-Bam*HI fragment containing the 3′ portion of the gene was inserted into the *Bam*HI site downstream of the U1 gene. The *Xba*I site at the junction is derived from a linker (see Figure 1).

BAL31 deletions

NS11Eco$^+$SalXba was opened either at the unique *Sal*I site or at the unique *Xba*I site, and treated with BAL31. The resected plasmids were then fitted with *Sal*I or *Xba*I linkers, recloned, and amplified in *E. coli*. The *Sal*I-*Xba*I fragments containing the 3′ portion of the U1 gene with selected deletions were isolated and recloned in pRec2 cleaved with *Sal*I and *Xba*I, to form the *Sal* and *Xba* series of mutant U1 genes. The precise endpoints of the BAL31 deletions were determined by DNA sequencing (Maxam and Gilbert, 1977).

Construction of the R series

To construct mutant genes of the R series, members of the *Xba* series were cleaved with *Xba*I. The ends were filled in with Klenow polymerase, and the DNA recut with *Bgl*II, which cleaves at the 5′ end of the U1 gene. The small fragments extending from the 5′ end of the U1 gene to the *Xba*I site at the endpoint of the deletions were isolated. Members of the *Sal* series were cleaved with *Sal*I, the ends were filled in with Klenow polymerase, and the DNA recut with *Bgl*II. The large fragments extending from the *Sal*I site through pBR322 and SV40 sequences to the *Bgl*II site at the 5′ end of the U1 gene were isolated, and ligated to the small fragments obtained from members of the *Xba* series. In the mutants so obtained, small segments of the U1 gene were replaced by a 12 nt sequence derived from the filled-in *Xba*I site ligated to the filled-in *Sal*I site. This reconstituted the *Sal*I site.

Transfections

HeLa cells were grown in Dulbecco's modified Eagle's medium supplemented with 5% foetal calf serum. DNA transfections were performed as described by Wigler *et al.* (1978) with some modifications: the DNA precipitate was allowed to form for only 5 to 10 min, and was left on the cells for 12 – 16 h. Cells were then shocked for 2 min with medium containing 20% dimethylsulfoxide (DMSO). RNA was extracted 44 to 50 h after transfection, as described by Berk *et al.* (1979).

Primer extension and S1 analysis

Primer and S1 probes were labeled by filling in with Klenow polymerase and [α-^{32}P]dATP (3000 Ci/mmol). Primer extension was performed as described (Hernandez and Keller, 1983), except that the hybridization temperature was 49°C. S1 mapping was performed as follows. Single-stranded DNA probe and RNA were mixed, lyophilized, and resuspended in 20 μl of hybridization buffer (80% formamide, 400 mM NaCl, 40 mM Pipes, pH 6.4, 1 mM EDTA). The mixture was heated for 2 min at 80°C, and incubated overnight at 49°C. 200 μl of S1 buffer (30 mM NaAc, pH 5.6, 3 mM $ZnSO_4$, 300 mM NaCl, 10 μg/ml denatured salmon sperm DNA) containing 500 to 1000 units of nuclease S1 were

then added, and incubation was continued for 2 h at 30°C. The nucleic acids were then precipitated by addition of 500 μl of ethanol, and loaded on a 6% sequencing gel.

Acknowledgements

I thank David Solnick, Manuel Ares, Marguerite Mangin, Elisabetta Ullu and Alan Weiner for help with the manuscript and numerous suggestions during the course of this work, and I thank Robert Ach for the compilation of DNA sequences. This work was supported by N.I.H. grants GM 31073 and GM 31335, and NSF grant PCM8315602 to Alan Weiner, in whose laboratory this work was performed. N.H. was supported by a postdoctoral fellowship from the Swiss National Fund, and more recently by postdoctoral fellowship DRG-851 from the Damon Runyon-Walter Winchell Cancer Fund.

References

Acheson,N.H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4754-4758.
Alonso,A., Jorcano,J.L., Beck,E., Hovemann,B., Schmidt,T. (1984a) *J. Mol. Biol.*, **180**, 825-836.
Alonso,A., Beck, E., Jorcano,J.L., Hovemann,B. (1984b) *Nucleic Acids Res.*, **12**, 9543-9550.
Amara,S.G., Evans,R.M., Rosenfeld,M.G. (1984) *Molec. Cell Biol.*, **4**, 2151-2160.
Ares,M., Mangin,M., Weiner,A.M. (1985) *Molec. Cell. Biol.*, in press.
Berget,S.M. (1984) *Nature*, **309**, 179-182.
Berk,A.J., Lee,F., Harrison,T., Williams,J., Sharp,P.A. (1979) *Cell*, **17**, 935-944.
Birchmeier,C., Folk,W., Birnstiel,M.L. (1983) *Cell*, **35**, 433-440.
Birchmeier,C., Grosschedl,R., Birnstiel,M.L. (1982) *Cell*, **28**, 739-745.
Birchmeier,C., Schümperli,D., Sconzo,G., Birnstiel,M.L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1057-1061.
Birnstiel,M.L., Büsslinger,M., Strub,K. (1985), *Cell*, in press.
Bringmann,P., Appel,B., Rinke,J., Reuter,R, Theissen,H., Lührmann,R. (1984) *EMBO J.*, **3**, 1357-1363.
Brown,D.T., Morris,G.F., Chodchoy,N., Sprecher,C., Marzluff,W.F. (1985) *Nucleic Acids Res.* **13**, 537-556.
Chandrasekharappa,S.C., Smith,J.H., Eliceiri,G.L. (1983) *J. Cell. Phys.*, **117**, 169-174.
Eliceiri,G.L. (1974) *Cell*, **3**, 11-14.
Eliceiri,G.L. (1980) *J. Cell. Phys.*, **102**, 199-207.
Eliceiri,G.L., Sayavedra,M.S. (1976) *Biochim. Biophys. Res. Commun.*, **72**, 507-512.
Fitzgerald,M., Shenk,T. (1981) *Cell*, **24**, 251-260.
Ford,J.P., Hsu,M.-T. (1978) *J. Virol.*, **28**, 795-801.
Frayne,E.G., Leys,E.J., Crouse,G.F., Hook,A.G., Kellems,R.E. (1984) *Molec. Cell. Biol.*, **4**, 2921-2924.
Frederiksen,S., Hellung-Larsen,P., Gram Jensen,E. (1978) *FEBS Lett.*, **87**, 227-231.
Georgiev,O., Birnstiel,M.L. (1985) *EMBO J.*, **4**, 481-489.
Gil,A., Proudfoot,N.J. (1984) *Nature*, **312**, 473-474.
Hagenbüchle,O., Wellauer,P.K., Cribbs,D.L., Schibler,U. (1984) *Cell*, **38**, 737-744.
Hashimoto,C., Steitz,J.A. (1984) *Nucleic Acids Res.*, **12**, 3283-3293.
Hentschel,C.C., Birnstiel,M.L. (1981) *Cell*, **25**, 301-313.
Hernandez,N., Keller,W. (1983) *Cell*, **35**, 89-99.
Higgs,D.R., Goodbourn,S.E.Y., Lamb,J., Clegg,J.B., Weatherall,D.J., Proudfoot,N.J. (1983) *Nature*, **306**, 398-400.
Hofer,E., Darnell,J.E.,Jr. (1981) *Cell*, **23**, 585-593.
Krämer,A., Keller,W., Apel,B., Lührmann,R. (1984) *Cell*, **38**, 299-307.
Krieg,P.A., Melton,D.A. (1984) *Nature*, **308**, 203-206.
Lerner,M.R., Steitz,J.A. (1981) *Cell*, **25**, 298-300.
Madore,S.T., Wieben,E.D. Pederson,T. (1984) *J. Cell. Biol.*, **98**, 188-192.
Manser,T., Gesteland,R.F. (1982) *Cell*, **29**, 257-264.
Marzluff,W.F., Brown,D.T., Lobo,S., Wang,S.-S. (1983) *Nucleic Acids Res.*, **11**, 6255-6270.
Mather,E.L., Nelson,K.J., Haimovich,J., Perry,R.P. (1984) *Cell*, **36**, 329-338.
Mattaj,I.W., Zeller,R. (1983) *EMBO J.*, **11**, 1883-1891.
Maxam,A.M., Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
McDevitt,M.A., Imperiale,M.J., Ali,H., Nevins,J.P. (1984) *Cell*, **37**, 993-999.
McLauchlan,J., Gaffney,D., Whitton,J.L., Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347-1368.
Montell,C., Fisher,E.F., Caruthers,M.H., Berk,A.J. (1983) *Nature*, **305**, 600-605.
Mount,S.M., Pettersson,I., Hinterberger,M., Karmas,A., Steitz,J.A. (1983) *Cell*, **33**, 509-518.
Murphy,J.T., Burgess,R.R., Dahlberg,J.E., Lund,E. (1982) *Cell*, **29**, 265-274.
Nevins,J.R., Darnell,J.E.,Jr. (1978) *Cell*, **15**, 1477-1493.
Nevins,J.R., Blanchard,J.-M., Darnell,J.E.,Jr. (1980) *J. Mol. Biol.*, **144**, 377-386.
Nojima,H., Kornberg,R.D. (1983) *J. Biol. Chem.*, **258**, 8151-8155.
Price,D.H., Parker,C.S. (1984) *Cell*, **38**, 423-429.
Roop,D.R., Kristo,P., Stumph,W.E., Tsai,M.J., O'Malley,B.W. (1981) *Cell*, **23**, 671-680.
Sheffery,M., Marks,P.A., Rifkind,R.A. (1984) *J. Mol. Biol.*, **172**, 417-436.
Simonsen,C.C., Levinson,A.D. (1983) *Molec. Cell. Biol.*, **3**, 2250-2258.
Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R., Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 8345-8352.
Stroke,I.L., Weiner,A.M. (1985) *J. Mol. Biol.*, in press.
Stunnenberg,H.G., Birnstiel,M.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6201-6204.
Strub,K., Galli,G., Büsslinger,M., Birnstiel,M.L. (1984) *EMBO J.*, **3**, 2801-2807.
Tani,T., Watanabe-Nagasu,N., Okada,N., Ohshima,Y. (1983) *J. Mol. Biol.*, **168**, 579-594.
Watanabe-Nagasu,N., Itoh,Y., Tani,T., Okano,K., Koga,N., Okada,N., Ohshima,Y. (1983) *Nucleic Acids Res.*, **11**, 1791-1801.
Weinberg,R., Penman,S. (1969) *Biochim. Biophys. Acta*, **190**, 10-29.
Westin,G., Lund,E., Murphy,J.T., Pettersson,U., Dahlberg,J.E. (1984) *EMBO J.*, **3**, 3295-3301.
Wickens,M., Stephenson,P. (1984) *Science*, **226**, 1045-1051.
Wigler,M., Pellicer,A., Silverstein,S., Axel,R. (1978) *Cell*, **14**, 725-731.
Yuo,C.-Y., Ares,M., Weiner,A.M. (1985) *Cell*, in press.
Zeller,R., Carri,M.-T., Mattaj,I.W., De Robertis,E.M. (1984), *EMBO J.*, **3**, 1075-1081.
Zhong,R., Roeder,R.G., Heintz,N. (1983) *Nucleic Acids Res.*, **11**, 7409-7425.
Zieve,G., Benecke,B.-J., Penman,S. (1977) *Biochemistry*, **16**, 4520-4525.
Zieve,G., Penman,S. (1976) *Cell*, **8**, 19-31.

Received on 22 April 1985