The EMBO Journal vol.8 no.12 pp.3875-3882, 1989

A U-snRNA gene-specific upstream element and a −30 'TATA box' are required for transcription of the U2 snRNA gene of *Arabidopsis thaliana*

Pierre Vankan and Witold Filipowicz

Friedrich Miescher-Institut, PO Box 2543, CH-4002 Basel, Switzerland

Communicated by H.J.Gross

The U2 and U5 snRNA genes of *Arabidopsis thaliana* contain in their promoter regions two elements with conserved sequence and position. To test the significance of this conservation we have made a construction in which the promoter of the U2 RNA gene is replaced by the synthetic 98 bp long sequence containing the two conserved elements: an upstream sequence element, GTCCCACATCG (USE, pos. −78 to −68), and a TATA-like sequence TATAAATA (−33 to −26), positioned approximately three helical turns apart, as in the wild-type promoter. This synthetic promoter efficiently drove transcription of the U2 gene in transfected protoplasts of *Nicotiana plumbaginifolia*. The importance of the individual elements and of their position within the promoter was investigated. Deletion of the USE, change of its orientation, and some single point mutations all decreased transcription 10- to 20-fold, and replacement of the TATA-like element by an unrelated sequence inactivated the promoter. Mutants in which the spacing between the USE and TATAAATA was changed were less active but no correlation was observed between promoter activity and insertion of either odd or even numbers of half helical turns. Insertion of a spacer between TATAAATA and the cap site resulted in accumulation of U2 RNA with an extended 5′ end, indicating that the TATAAATA element is responsible for selection of the initiation site. The data indicate that the promoters of RNA polymerase II-specific U-snRNA genes in higher plants differ from their animal counterparts and also from plant mRNA gene promoters. They contain two essential elements, an USE, an element found only in U-snRNA genes, and a TATA element which is indistinguishable from the TATA boxes of mRNA-coding genes.

Key words: plant genes/promoter structure/transcriptional signals/transient expression in protoplasts/U-snRNA genes

Introduction

All eukaryotic cells contain in their nuclei a group of abundant small RNAs, the U-snRNAs, which are components of ribonucleoprotein particles, the U-snRNPs. Some of the U-snRNPs (U1, U2, U5 and U4/U6) play a role in pre-mRNA splicing (for reviews, see Maniatis and Reed, 1987; Guthrie and Patterson, 1988). U1, U2 and U5 snRNPs are involved in recognition of the 5′ splice site, the branch site and probably the 3′ splice site respectively. The function of U4 and U6 snRNPs, which are found together in one particle, is unknown. In addition to U1−U6 snRNAs, several other U-RNAs have been identified. Some of them are involved in RNA processing reactions other than splicing (reviewed by Birnstiel and Schaufele, 1988; Reddy and Busch, 1988).

The synthesis of U-snRNAs has been extensively studied in vertebrate cells (for reviews, see Dahlberg and Lund, 1988; Parry *et al.*, 1989a). All U-snRNA genes, with the exception of the U6 gene, are transcribed by RNA polymerase II, since their transcription is sensitive to α-amanitin (Gram-Jensen *et al.*, 1979) and the primary transcripts contain a 7-methylguanosine (m^7G) cap. The m^7G is further modified in the cytoplasm to a 2,2,7-trimethylguanosine structure (Mattaj, 1986). However, several features distinguish U-snRNA synthesis from mRNA gene transcription. In contrast to mRNA coding genes, the promoters of U-snRNA genes do not contain a TATA box at position −30, but rather a U-snRNA gene-specific proximal sequence element (PSE) in the −60/−50 region, with a sequence unrelated to a TATA box (Skuzeski *et al.*, 1984; Ares *et al.*, 1985; Ciliberto *et al.*, 1985; Mattaj *et al.*, 1985; Murphy *et al.*, 1987; Parry *et al.*, 1989b; see also Dahlberg and Lund, 1988; Parry *et al.*, 1989a). This element is a functional equivalent of the TATA box as it is required for selection of the correct start site by the RNA polymerase II complex. In addition, an enhancer-like distal sequence element (DSE) is present in the −250 to −200 region of U-RNA promoters. It is usually composed of several sequence motifs, such as octamer- or SP1-binding sites, similar to the ones found in mRNA genes (Skuzeski *et al.*, 1984; Westin *et al.*, 1984; Ares *et al.*, 1985, 1987; Ciliberto *et al.*, 1985; Krol *et al.*, 1985; Mattaj *et al.*, 1985; Mangin *et al.*, 1986; Kazmaier *et al.*, 1987; Janson *et al.*, 1987). Consistent with this similarity is the finding that the mRNA gene enhancer of SV40, containing octamer-like motifs, can efficiently activate transcription of a human U2 gene devoid of its own enhancer (Mangin *et al.*, 1986). However, the DSE from the *Xenopus* or human U1 genes cannot, when tested in *Xenopus* oocytes or in human 293 cells, replace the upstream element present in an mRNA coding gene (Ciliberto *et al.*, 1987; Dahlberg and Schenborn, 1988). Tanaka *et al.* (1988) have recently demonstrated that in the context of the human U2 gene promoter, the octamer motif behaves differently from the octamer motif found in mRNA-coding genes. It is likely that the nature of the protein factor which interacts with the octamer sequence is determined by the prior binding of distinct transcription factors to the proximal sequence elements, the TATA box present in mRNA genes and the −60 to −50 PSE sequence present in U-RNA genes.

U-snRNAs are not polyadenylated. In vertebrates their genes contain in the 3′ non-coding region a 12−15 bp long sequence which is required for transcription termination or 3′ end processing (reviewed by Dahlberg and Lund, 1988). Interestingly, appropriate termination of U-RNA synthesis requires that transcription is initiated from a U-RNA gene

promoter. When transcription of a U-RNA gene is driven by an mRNA promoter, the 3′ end signal is not used and RNA transcripts extend to a polyadenylation signal localized further downstream (Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986; Hernandez and Lucito, 1988). These results further underscore the differences between U-RNA and mRNA gene transcription and suggest that distinct RNA polymerase II complexes are involved in synthesis of the two classes of RNAs.

We have recently characterized several active genes encoding U2 and U5 snRNAs in the higher plant *Arabidopsis thaliana* (Vankan and Filipowicz, 1988; Vankan *et al.*, 1988). The upstream regions of these genes contain two conserved sequence elements. One of them, having a consensus TATAAATA and located about position −30, is similar to the TATA boxes of mRNA coding genes (Joshi, 1987). The second is USE, RTCCCACATCG located in the −70 region; so far this element has not been encountered in mRNA gene promoters. Similar upstream sequences are also present in the U1 RNA genes of bean and soybean, although activity of these genes has not been demonstrated (van Santen and Spritz, 1987; van Santen *et al.*, 1988). An additional interesting feature of the U-RNA gene promoters in higher plants is the strong conservation of spacing between the two upstream sequence elements. The USE and the TATA-like sequence are separated by approximately three helical turns of DNA in all genes characterized so far. Immediately adjacent to the 3′ end of the coding regions of plant U1, U2 and U5 snRNA genes is the sequence $CAN_{4-9}AGTN(A/T)AA$ which may represent a transcription termination or RNA processing signal (Vankan *et al.*, 1988).

The unique features of the upstream non-coding regions of plant U-RNA genes prompted us to investigate their transcription in more detail. We have constructed a model gene in which the natural promoter of the *Arabidopsis* U2 gene is replaced by a synthetic sequence containing the conserved elements in the appropriate positions. This model gene was used to test the importance of the conserved sequence elements and their spacing for gene transcription in transfected plant protoplasts.

Results

Construction of a synthetic promoter

Comparison of the upstream non-coding regions in plant U-snRNA genes revealed the presence of two elements with conserved sequence and position: the sequence TATAAATA in the −30 region and the USE of sequence RTCCCACATCG around position −70 (Vankan *et al.*, 1988). Deletion analysis, carried out with one of the *Arabidopsis* U2 RNA genes (gene U2.2), has indicated that sequences further upstream are probably not important for transcription in transfected plant protoplasts; promoter deletions to positions −244 or −148 decreased activity by only 20−30%, while deletion to −44, removing the USE, decreased transcription >10-fold (Vankan and Filipowicz, 1988).

In order to analyse the structure of the *Arabidopsis* U2 promoter in more detail we have constructed a model gene, called U2.SP, in which the original 388 bp long 5′ flanking region of the *Arabidopsis* U2.2 gene is replaced by a 98 bp long synthetic sequence, while the coding and 3′ non-coding regions remain unchanged. The assembly of the model gene is schematically presented in Figure 1 and the sequence of the synthetic promoter is shown in Figure 2, together with the sequences of the *Arabidopsis* U2 and U5 gene promoters. The synthetic promoter contains the USE element GTCCCACATCG and the sequence TATAAATA at positions −78 to −68 and −33 to −26, exactly as in the natural U2.2 gene promoter. In addition, several other short nucleotide stretches, partially conserved between *Arabidopsis* U2 and U5 genes but not among other plant U-RNA genes (Figure 2, and Vankan *et al.*, 1988), were incorporated into the original version of the synthetic promoter. These include: the sequence AAA located 5′ of the USE (pos. −81 to −79); the sequence GTAG located 5′ of the TATAAATA box (pos. −37 to −34); and the sequence CAATTC (pos. −6 to −1) located upstream of the cap site. In the latter sequence only the C at position −1 is conserved among all U1, U2 and U5 RNA genes isolated from plants. The remaining sequences incorporated into the synthetic promoter differ from the original U2.2 sequence and contain several restriction sites to facilitate manipulation of DNA (Figures 1 and 2).

Activity of the U2.SP gene

The activity of the U2.SP gene containing the synthetic promoter was compared with the wild-type U2.2 gene. The latter gene, containing 388 bp of the authentic upstream non-coding sequence, was previously found to be the most efficiently expressed among three different *Arabidopsis* U2 genes tested in a transient expression system (Vankan and Filipowicz, 1988). In the experiment shown in Figure 3, increasing concentrations of plasmids pU2.SP and pU2.2 were used for transfection of *Nicotiana plumbaginifolia* protoplasts. The level of transcription was measured by RNAse A/T_1 mapping, using a ^{32}P-labelled RNA probe complementary to the U2.2 plus strand. Maximal yield of U2 RNA was obtained at 25−50 μg of either plasmid used for transfection. Within the linear range of response, the model U2.SP gene was transcribed at ~65−70% of the efficiency of the wild-type gene. The conserved sequences incorporated into the synthetic promoter therefore contain sufficient information for correct and efficient initiation of transcription of the U2 RNA gene in protoplasts.

Mutations in the upstream sequence element

We have used the U2.SP gene to assess the importance of individual conserved sequence elements incorporated into the synthetic promoter. First, we have tested a series of constructs containing mutations in the USE region. Deletion of the USE box (mutant U2.SPΔUSE) resulted in a 20- to 40-fold decrease of activity (Figures 4 and 6); residual transcription yielded RNA of a size identical to the RNA produced from the control plasmid. In the mutant U2.SP-CCC the sequence AAA (pos. −81 to −79), directly 5′ of the USE, was replaced with CCC. The sequence AAA is only conserved among the *Arabidopsis* U2 RNA genes. This mutation had no effect on transcription (Figure 4, lane 1).

Single point mutations (transversions) were generated in all positions of the USE, except for the positions T_2 and A_6, which were obtained as double-point mutants $T_2C_5 \rightarrow G_2A_5$ and $A_6T_9 \rightarrow C_6G_9$. The $G_1 \rightarrow T_1$ and the $C_3 \rightarrow A_3$ mutations had no significant effect on transcription. For the $G_1 \rightarrow T_1$ mutant this is consistent with the finding that this base is only conserved between the U2 genes of *Arabidopsis*.

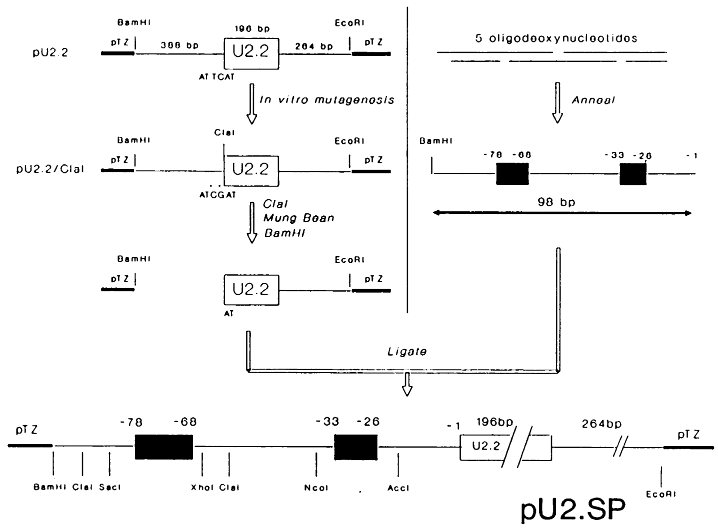

Fig. 1. Construction of the plasmid pU2.SP containing the U2 RNA gene with a synthetic promoter. Removal of the wild-type promoter from the *Arabidopsis* gene U2.2 is indicated on the left. Mutations introduced into the pU2.2 plasmid to generate the pU2.2/*Cla*I derivative are marked by asterisks. Digestion of pU2.2/*Cla*I with *Cla*I and *Bam*HI removed 388 bp of upstream sequence. After restriction with *Cla*I and blunting the DNA end corresponds exactly to the 5′ terminus of the coding region (AT...; Vankan and Filipowicz, 1988). Assembly of the synthetic promoter by annealing of five oligodeoxynucleotides (indicated by five thin lines) is shown on the right. The length of the annealed fragment is indicated by the double-pointed thick arrow. Two conserved elements, the USE (pos. −78 to −68) and the TATAAATA box (pos. −33 to −26) present in the synthetic sequence are drawn as black boxes. The resulting construction pU2.SP is shown at the bottom of the scheme. The 196 bp long coding sequence of the U2.2 gene is represented by an open box. Lengths of the 5′ and 3′ non-coding sequences and relevant restriction sites are indicated. pTZ vector sequences are represented by thick lines.

```
                        -80                 -60                 -40                 -20                   -1
                .         .         .         .         .         .         .         .         .         .
U2    nnnnnnnnnnnnnntgnnnaaaGTCCCACATCGnnAAgnnanngAnnnntnnnnAnAAnnnnnGTagTATAaatAnnaagcnnnGnCnn--tcgnncAAnTC
                           | ||||||||||  |||  |        |              |||||||||||     |     |      |   || ||
U5    TTTAACAAGGGAACTTCAAGTATCCCACATCGGAAAGCTAAGAGACGGTGATAGAGGAAATGATAGTATAAATAAACTTCACTCTCCTGGGAGTAAAAATC

SP    GGATCCATCGATGAGCTCAAAGTCCCACATCGGGCTCGAGCATCGATACCTGAGGCTCCATGGTAGTATAAATAACTGGTATACTGGT--GACCTCAATTC
```

Fig. 2. Alignment of the sequences representing a consensus derived from the promoters of six *Arabidopsis* U2 RNA genes (**upper lane**), the *Arabidopsis* U5 gene promoter (**middle lane**) and the synthetic promoter (SP, **lower lane**). Nucleotides of the U2 consensus that are conserved in six out of six and five out of six genes are shown in upper-case and lower-case letters respectively. Nucleotides conserved between the U2 consensus and U5 gene are indicated by vertical lines. Sequences corresponding to the USE (pos. −78 to −68) and TATAAATA (−33 to −26), highly conserved among all sequenced plant U-RNA genes (see Vankan *et al.*, 1988) are underlined with double lines, and the sequences partially conserved only in *Arabidopsis* genes are underlined with a single line. Numbering corresponds to the U2 consensus and the synthetic sequence.

In the *Arabidopsis* U5 gene (Vankan *et al.*, 1988), and the U1 genes of soybean and bean (van Santen and Spritz, 1987; van Santen *et al.*, 1988) an A is found in this position. For the $C_3 \rightarrow A_3$ mutation, the lack of effect is surprising, as this base is conserved in all plant U-RNA genes sequenced so far. Point mutations in the USE positions 5, 7, 8, 10 and 11 reduced transcription 4- to 10-fold, whereas mutations in the positions 4 and 9 lowered transcription to a level comparable to the mutant from which the USE was deleted. The effect of mutation at position A_6 cannot be assessed since the double-point mutant $A_6T_9 \rightarrow C_6G_9$ had the same low activity as the point mutant $T_9 \rightarrow G_9$. Mutant $T_2C_5 \rightarrow G_2A_5$ was, however, significantly less active than the mutant $C_5 \rightarrow A_5$, indicating that the conserved T residue at position 2 contributes to the function of the element.

Mutations in the TATA-like box region

Several mutants were constructed in order to test the importance of the conserved nucleotides around position −30 and to determine whether the TATA-like element present in this region is involved in determination of the transcription initiation site (Figure 5). In the mutant U2.SP-TATA1 four bases positioned upstream of the TATA-like element were changed from GTAG to TGCT. The sequence RTAG is conserved in the *Arabidopsis* U5 and U2 genes (see Figure 2, and Vankan *et al.*, 1988), but not in the U1 genes of bean and soybean. This mutation did not affect transcription efficiency (Figure 5A, lane 3). In construction pU2.SP-TATA2 the conserved element TATAAATA (pos. −33 to −26), was replaced by the sequence GTGCACGA. This mutation completely eliminated expression of U2 RNA (lane 4). In construction U2.SP-TATA3 the element was replaced by the TATA-like motif TTTATATA, which is conserved in the −30 region of three different U6 snRNA genes of *Arabidopsis*; these genes are transcribed by RNA polymerase III (F.Waibel and W.Filipowicz, manuscript in preparation). The change had no effect on expression of the U2 gene (lane 5), indicating that either the same transacting factor is involved in transcription of U2 and U6 RNA genes or that a hypothetical U2 gene-specific factor can tolerate two point mutations.

We have tested whether an increase in distance between the TATAAATA sequence and the normal transcription start site would lead to the production of U2 RNA with an extended 5′ end. Such a result would be consistent with a role of TATAAATA element in specifying the transcription

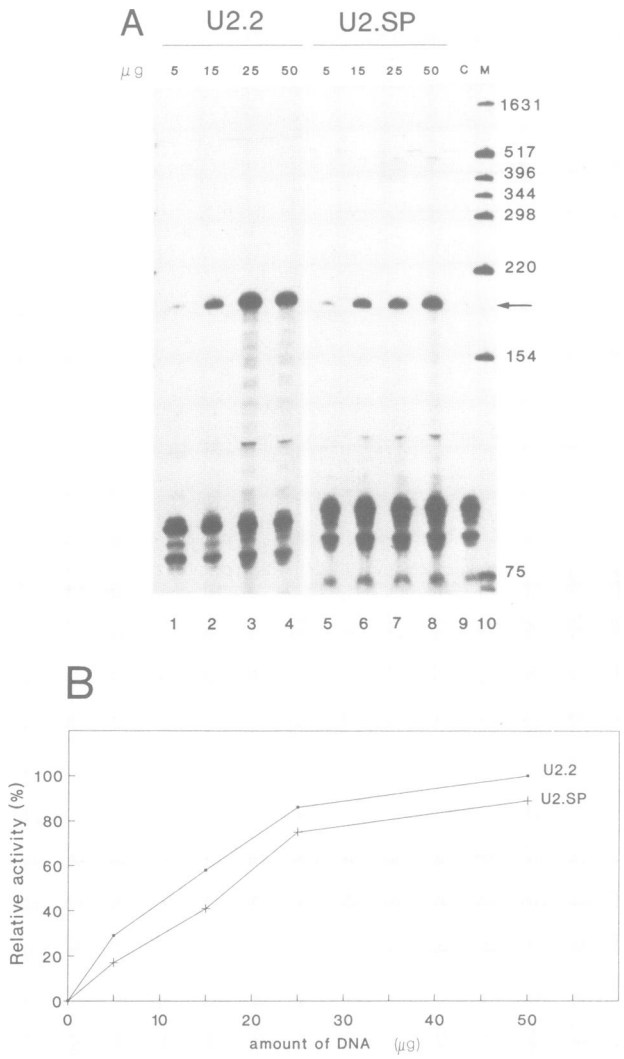

Fig. 3. Comparison of the activities of U2 RNA genes with natural and synthetic promoters. RNAs isolated from *N.plumbaginifolia* protoplasts transfected with increasing amounts of either pU2.2 (**lanes 1–4**) or pU2.SP (**lanes 5–8**) were analysed by RNase mapping. (**A**) Autoradiogram of a polyacrylamide gel from one representative experiment. U2 RNA length fragments are indicated by arrow. Amounts of DNA used for transfection are shown at the top. **Lane 9**, mapping of RNA form mock-transfected protoplasts. **Lane 10**, size markers (3′ end-labelled *Hin*fI digest of pBR322). (**B**) Protected fragments corresponding in size to full-length U2 RNA were quantitated by scintillation counting. Values were normalized using the endogenous U2-RNA-protected fragments (80–90 nt bands seen in panel A) as reference and plotted as percentage of the maximal wild-type U2.2 gene activity. Each point represents an average of three independent transfections.

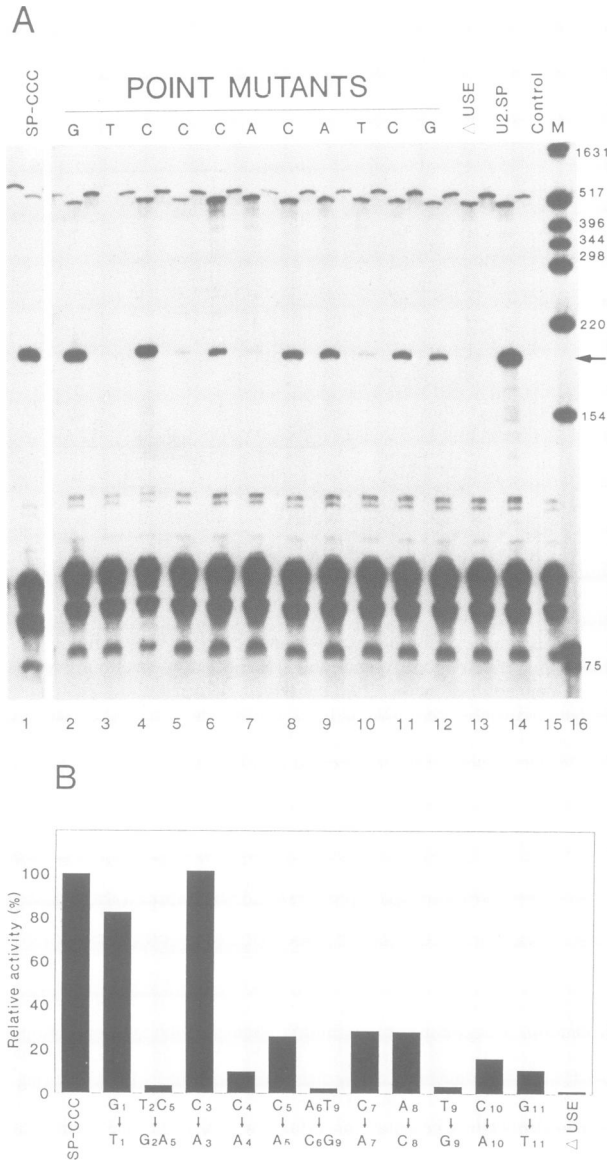

Fig. 4. Activity of the U2.SP genes containing mutations in the USE region. (**A**) Autoradiogram of a gel representing RNase mapping. Mutants used for protoplast transfection are described in Materials and methods, and in Results. Point mutations (**lanes 2–12**), the nature of which is shown in (B), are arranged according to their position in the USE, the sequence of which is shown at the top. Note that no single point mutations were obtained for positions T_2 and A_6. **Lanes 14** and **15**, protoplasts transfected with non-mutated gene U2.SP and mock-transfected respectively. **Lane 16**, size markers. (**B**) Quantitation of the expression data for the USE mutants. A value of 100% corresponds to the activity of a non-mutated U2.SP gene. Each value is an average of two independent transfections.

initiation site. Indeed, expression of the U2.SP-TATA4 mutant, containing a 22 bp insertion downstream of the TATAAATA, resulted in accumulation of RNA which protected an RNA probe fragment 6–8 nt longer than the wild type U2 RNA. Since complementarity of the U2.2 antisense probe used for RNase mapping extended only 7 nt upstream of the 5′ end of the U2 RNA (Figure 5B; Vankan and Filipowicz, 1988), we have used another probe, complementary to the U2.SP gene, in order to locate the site of initiation more precisely. The U2.SP gene-specific probe yielded a protected fragment ~20 nt longer than authentic U2 RNA (data not shown). Hence, an insertion of 22 bp between the TATAAATA element and the cap site resulted in the upstream shift of the transcription start by similar distance. The observation that transcription can be initiated upstream of the physiological cap site also indicates that the nucleotides CAATTC (pos. −6 to −1), which are partially conserved among the *Arabidopsis* U5 and U2 snRNA genes, but not among bean and soybean U1 genes, are not required for initiation of transcription. However, it should be noted that the amount of RNA transcribed from pU2.SP-TATA4 was 3- to 4-fold lower than the amount of RNA transcribed from pU2.SP. This could be due to either a lower rate of transcription or decreased stability of the longer U2 RNA transcripts *in vivo*.

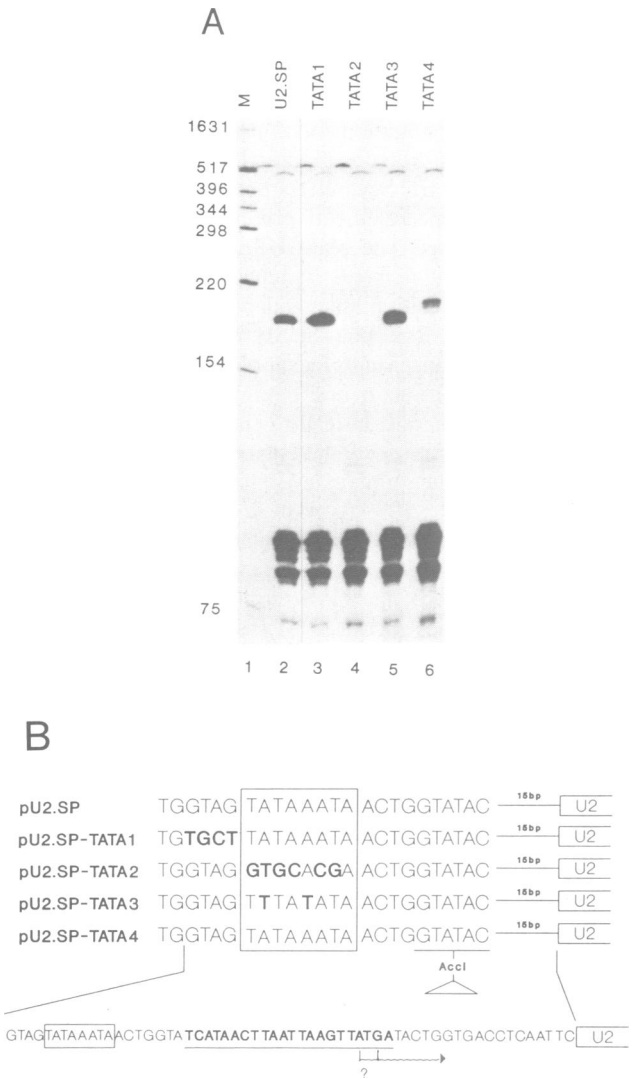

Fig. 5. Activity of the U2.SP genes with mutations in the TATA region. **(A) Lane 2**, mapping of RNA from protoplasts transfected with pU2.SP. **Lanes 3–6**, analysis of RNAs from protoplasts transfected with plasmids pU2.SP-TATA1, -TATA2, -TATA3 and -TATA4 respectively. **Lane 1**, size markers. **(B)** Sequences of the modified TATA box regions in plasmids pU2.SP-TATA1, TATA2, TATA3 and TATA4. The sequence of pU2.SP is also shown. The conserved TATA box sequence is marked with a box. Nucleotide changes are in bold letters. The upstream non-coding sequence of pU2.SP-TATA4, a mutant with 22 bp insertion (underlined bold letters) into the *Acc*I site is shown at the bottom. Putative transcription initiation sites are indicated.

Changes of spacing between the USE and TATA-like element

The highly conserved 33–34 bp spacing between the USE and the TATA element in the U1, U2 and U5 promoters is an interesting feature of the U-snRNA genes in plants (Vankan *et al.*, 1988). In order to test the importance of this conservation we have constructed a series of the U2.SPace mutants, in which the distance between the two sequence elements is altered (Figure 6C). In mutants −0.5, −1 and −2 the distance is shortened by 5, 10 and 20 bp, which corresponds to approximately one-half, one and two helical turns of DNA respectively. In mutants +0.5 through +5 the distance is increased by the indicated number of helical turns. Shortening of the distance between the USE and the TATA box by one or two helical turns decreased

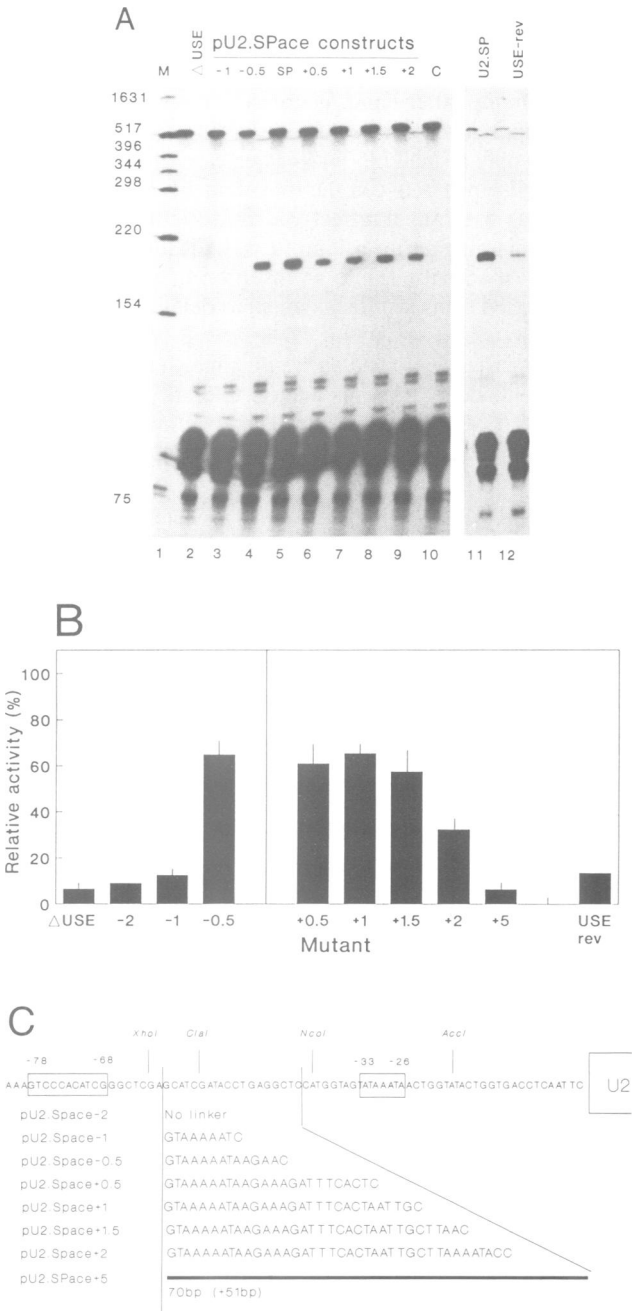

Fig. 6. Activity of the U2.SP genes with modified spacing between the USE and TATA elements, and of the U2.SP-USErev which contains the USE in reverse orientation. **(A)** Autoradiogram of a gel from one representative experiment. Mutants used for transfection are indicated at the top and described in (C) or in Materials and methods (USErev). Changes in number of helical turns between the USE and TATA elements in the U2.SPace mutants are indicated below the line, with SP corresponding to the non-mutated U2.SP gene (**lanes 3–9**). Analysis of the −2 and +5 mutants is not shown. **Lane 10**, mock-transfected protoplasts; **lanes 11** and **12**, protoplasts transfected with pU2.SP and pU2.SP-USErev respectively; **lane 1**, size markers. **(B)** Quantitation of the activities of different mutants. Activity expressed is relative to the activity of unmodified pU2.SP. Values represent means ± SEM from 4–6 independent transfections, with an exception of −2, +5 and USErev mutants which were tested one, three and two times respectively. **(C)** Structure of different U2.SPace mutants. Sequences inserted between the *Xho*I and *Nco*I sites of the synthetic promoter are shown. The insert in pU2.SPace+5 is GCGCTCGTACCTGCATGATATCATAAAACCATGGAAAATGTT-AGCCAAAAAAATGGAAAAGAAGAAATAA. In this construct nucleotides CATGGTA positioned downstream of the insertion are replaced by ATAATTT.

transcription of the U2 gene 10- to 15-fold (Figure 6). The activity of these mutants was close to the activity of the construct pU2.SPΔUSE from which the USE had been deleted. On the other hand, decreasing the spacing by half a turn or increasing it by half, one or one and a half turns resulted in only a small (35−45%) decrease of activity. Only when the distance between the two elements was increased further was a stronger effect on transcription seen. The mutant containing an insertion of two helical turns retained only 32% of activity, while an insertion of five additional turns inhibited transcription as much as deletion of the USE (Figure 6).

The properties described above distinguish the USE from classical enhancer elements which tolerate changes in a location and also function in an orientation-independent manner (reviewed by Wasylyk, 1988). In order to test whether activity of the USE depends on its orientation, we have constructed the mutant U2.SP-USErev in which USE is located in a similar position as in a wild-type gene but the orientation of the element is reversed. This mutant was transcribed at 12% efficiency, only about twice as efficiently as the mutant without USE (Figure 6).

Discussion

The results of this work indicate that two promoter elements, a USE with sequence RTCCCACATCG and a TATA-like element TATAAATA located in the −70 and −30 regions of plant U-snRNA genes, are necessary and sufficient for transcription of the *Arabidopsis* U2 RNA gene in transfected protoplasts. Our approach to investigate the requirements of U-snRNA gene transcription in plants was to construct the U2.SP model gene in which the natural promoter of the *Arabidopsis* U2 gene is replaced by a synthetic promoter of mostly arbitrary sequences but which contains the elements that are conserved between different *Arabidopsis* U2 and U5 RNA genes. The activity of the U2.SP gene and its variants which have the sequence or the position of the individual elements modified was investigated in protoplasts of *N.plumbaginifolia*. The model U2.SP gene was transcribed at 65−70% efficiency of the wild-type U2.2 gene. We have shown previously that deletion of the upstream sequences to position −148 from the wild-type gene U2.2 also resulted in ~30% decrease of activity (Vankan and Filipowicz, 1988). This moderately lower activity of the U2.SP gene and the U2.2 deletion mutant −148 are likely to be caused by the alien, relatively GC-rich plasmid sequences brought to the proximity of the promoter. The upstream sequences of the *Arabidopsis* U2 and U5 RNA genes are generally AT-rich, which may be of importance for optimal promoter function. Nevertheless, we cannot exclude the possibility that far upstream regions of the U2 gene promoters contain some other sequence motifs which would have a small modulating effect on gene transcription. Lack of sequence conservation in that region, however, argues against this possibility. It has been verified that transcription of the U2.2 and U2.SP genes in transfected protoplasts has the sensitivity to α-amanitin expected for RNA polymerase II genes (Mühlbach and Saenger, 1979). α-Amanitin at a concentration of 50 μg/ml strongly inhibits the expression of transfected U2 RNA and mRNA genes and has no effect on expression of U6 snRNA genes which are transcribed by RNA polymerase III (F.Waibel and W.Filipowicz, unpublished results).

Mutational analysis of the USE has shown that this element has a strong stimulatory effect on transcription. The low amounts of U2 RNA transcribed in the absence of USE still had correct size indicating that this element is not responsible for location of the initiation site. Similarly, mutations in the USE and changes in its location had no effect on selection of transcription start. Several features distinguish the USE from distal sequence elements present in U-snRNA genes of other organisms. The plant USE has no properties of classical enhancer elements which usually function in an orientation- and position-independent fashion (reviewed by Wasylyk, 1988). The plant USE is only active in one orientation and changes in its location markedly affect activity of the U2 gene promoter (see below). This contrasts with the DSE of vertebrate U-RNA genes which functionally resemble in many respects the enhancer elements found in protein coding genes (for reviews, see Dahlberg and Lund, 1988; Parry *et al.*, 1989a). Also in contrast to vertebrate DSE motifs, the sequences of which are identical to enhancer elements found in mRNA genes (Dahlberg and Lund, 1988; Parry *et al.*, 1989a), the USE motif has only been found so far in plant genes encoding U1 (van Santen *et al.*, 1988; T.Kiss and F.Solymosy, personal communication), U2 (Vankan and Filipowicz, 1988), U3 (T.Kiss and F.Solymosy, personal communication), U5 (Vankan *et al.*, 1988) and U6 (F.Waibel and W.Filipowicz, manuscript in preparation) RNAs. Analysis of the upstream sequences of protein-coding genes of higher plants available in the EMBL and GenBank data libraries did not reveal elements similar to the USE.

Replacement of the TATAAATA in the −30 region by an unrelated sequence completely eliminated activity of the U2.SP gene. This result is consistent with the strong conservation of this element, both in respect to sequence and position, in all plant U-RNA genes characterized to date (Vankan *et al.*, 1988; T.Kiss and F.Solymosy, personal communication; F.Waibel and W.Filipowicz, manuscript in preparation). The TATAAATA element appears to be responsible for selection of the transcription initiation site as an insertion of a 22 bp spacer between the TATAAATA sequence and the cap site of the U2.SP gene resulted in a upstream shift of the transcription start by a similar distance. Both the consensus sequence TATAAATA and the location in the −30 region make the U-RNA gene TATA element indistinguishable from the TATA boxes present in plant mRNA coding genes (Joshi, 1987). This strongly contrasts with the situation in vertebrate U-RNA genes. In these genes another element, positioned in the −50 to −60 region, with a sequence unrelated to the TATA box of mRNA genes, is responsible for selection of the start site by RNA polymerase II (for reviews, see Dahlberg and Lund, 1988; Parry *et al.*, 1989a).

One of the unique features of the U1, U2 and U5 gene promoters in plants is a strong conservation of the 33−34 bp distance between the USE and the TATA box (Vankan *et al.*, 1988). The two sequence elements are therefore centred approximately four DNA helical turns apart, suggesting that transcription factors interacting with them have to bind to the same side of the helix. The requirement for different transcription factor binding sites being positioned on the same side of the DNA has been previously demonstrated for the SV40 early promoter (Takahashi *et al.*, 1986). It was therefore surprising to find that, despite the strong conservation of location of the USE and TATA elements,

the spacing between them can be changed considerably without strongly affecting activity of the promoter (mutants U2.SPace−0.5 through +2). Deletion of half a turn or insertion of a half or one and a half turns did not inhibit transcription more strongly than insertion of one or two whole helical turns of DNA. These results indicate that the USE and TATA-like elements do not have to be positioned on the same side of the helix in order to activate transcription. Proximity between them is essential, however, since increasing the distance between the elements to five helical turns eliminated most of the activity. Similar observations were made with the adenovirus 2 E1B promoter which contains TATA and the Sp1 elements separated by 8 bp; Wu and Berk (1988) have found that the spatial relationship between these two elements is quite flexible but an increased distance between them to 24 or 30 bp is not tolerated. It will be of interest to test the activity of different U2.SPace mutants in stably transformed plants. It is possible that within a chromatin context the requirement for appropriate spacing between the USE and TATA elements is more stringent.

The strong 10- to 15-fold decrease of activity observed with the U2.SPace mutants −1 and −2 is of special interest. We have recently established that the promoters of the *Arabidopsis* U6 RNA genes, which are transcribed by the RNA polymerase III, contain both the USE and the −30 TATA-like elements. The only apparent feature which distinguishes the U6 gene promoter from the RNA polymerase II-specific promoters is a one helical turn closer spacing between the USE and the TATA box (F.Waibel and W.Filipowicz, manuscript in preparation). The construct pU2.SPace−1 therefore strongly resembles an RNA polymerase III-specific promoter. If RNA synthesis from this construction is indeed initiated by RNA polymerase III, transcription would terminate in the coding region of the U2 gene, where several stretches of T residues, which usually function as RNA polymerase III termination signals (Bogenhagen and Brown, 1981), are present. Alternatively, a low activity of the U2.SPace−1 and −2 mutants could be explained by too close proximity of the USE and TATA boxes which might prevent simultaneous binding of trans-acting factors.

Clearly, in both vertebrates and higher plants, the promoters of U-RNA genes transcribed by RNA polymerase II have unique structural features which distinguish them from the promoters of protein-coding genes. A distinct organization of U-RNA gene promoters must be essential for securing highly efficient and constitutive synthesis of U-snRNAs (reviewed by Reddy and Busch, 1988), large amounts of which are required for the processing of multiple introns from vertebrate and plant pre-mRNAs. The organization of the U-snRNA gene promoters is, however, diametrically different in plant and vertebrate cells. In vertebrates it is the proximal PSE sequence which distinguishes U-RNA genes from protein-coding genes. In higher plants U-RNA genes differ from mRNA genes by the presence of a specific distal element RTCCCACATCG which functions in conjunction with a proximal −30 TATA element found in both types of genes.

Some data about transcription of the RNA polymerase II-encoded U-snRNA genes in sea urchin and yeast has also become recently available. The U7 snRNA genes in sea urchin *Psammechinus miliaris* contain a −63 to −26 proximal promoter element comprising two segments, located around positions −50 and −30, which interact with protein factors. The −55 to −44 binding site is indispensable for transcription *in vitro*; it resembles a TATA motif but is situated at the position of the PSE element of vertebrate U-RNA genes rather than at the −30 region of mRNA genes. Sequences upstream of position −80 enhance transcription of sea urchin genes ~5-fold *in vivo* but not *in vitro*; sequences responsible for this activation have not yet been identified (Southgate and Busslinger, 1989).

It appears that promoters of U-snRNA and mRNA genes in the yeast *Saccharomyces cerevisiae* are much more related to each other than in vertebrates and plants. Although the sequences which direct U-snRNA gene transcription in yeast have not yet been characterized, a comparison of the upstream regions in U-snRNA genes has revealed the presence of a TA-rich element which is similar, but not identical, to the yeast mRNA TATA box. The consensus of U-snRNA genes is TATAAARRGNR as opposed to the shorter consensus TATAAA found in mRNA genes (Hahn *et al.*, 1985; Parker *et al.*, 1988). Furthermore, the distance between the TATA-like box and the cap site in U-snRNA genes is more tightly constrained than in mRNA genes. The U-snRNA TATA box is localized between positions −85 and −101 as opposed to the positions −40 to −120 in mRNA promoters (Hahn *et al.*, 1985; Parker *et al.*, 1988). Variation in the position of the TATA box and the frequent presence of multiple TATA elements in mRNA promoters make yeast mRNA transcripts highly heterogenous at their 5′ ends (Hahn *et al.*, 1985; Hinnebusch and Fink, 1983); in contrast, the initiation of snRNA gene transcription is highly accurate. Another feature that distinguishes transcription of U-snRNA genes in yeast from vertebrates is an apparent lack of coupling between initiation and termination events. Patterson and Guthrie (1987) have demonstrated that transcription of the yeast U5 RNA gene driven by the GAL1 promoter generates functional snRNA molecules with correct 3′ ends. We are presently studying whether initiation and termination processes are coupled during transcription of plant U-snRNA genes.

Materials and methods

Construction of pU2.SP and its derivatives

Unless indicated otherwise, all techniques used for manipulating DNA were as described in Maniatis *et al.* (1982). The identity of all constructions was verified by sequence analysis using the dideoxynucleotide method and double-stranded DNA.

Parent construct pU2.SP

The construction is outlined in Figure 1. Plasmid pU2.SP is a derivative of pU2.2 which contains a 0.85 kb *Bam*HI−*Eco*RI fragment of the *Arabidopsis* U2.2 gene (Vankan *et al.*, 1988) cloned into the pTZ18R vector (United States Biochemical Corp.). pU2.2/*Cla*I, containing a *Cla*I site on the 5′ border of the U2.2 gene coding region, was prepared from pU2.2 by site-directed mutagenesis (Zoller and Smith, 1983) using single-stranded plasmid DNA, a mutant oligonucleotide primer and T4 DNA polymerase. The *Cla*I site was introduced in such a way that after cutting and blunting only the coding bases of the U2.2 gene remain. In order to remove the natural 5′ non-coding region, plasmid pU2.2 was digested sequentially with *Cla*I, mung bean nuclease and *Bam*HI. Five overlapping oligodeoxynucleotides, 25−50 nt in length, were annealed to generate the 98 bp long synthetic promoter. The annealing reaction (100 μl) contained 20 mM Tris−HCl, pH 7.5, 20 mM NaCl and 100 pmol of each oligonucleotide. The solution was heated to 80°C and was then slowly cooled down to 20°C. To obtain pU2.SP, the annealed fragment with proper *Bam*HI and blunt ends was ligated into pU2.2/*Cla*I, restricted as described above.

Mutants in the USE region

Most of the point mutants in the USE were generated by exchanging the *Sac*I−*Xho*I fragment of the synthetic promoter in pU2.SP with mutated

sequences, using the mixed linker insertion method (Hutchison *et al.*, 1986). During synthesis of the plus strand oligonucleotide 1:9 mixtures of T:G, G:T, A:C and C:A were used where G, T, C and A respectively normally occur in the USE region of the sequence. The complementary oligodeoxynucleotide was synthesized in an analogous way. Oligonucleotides were annealed in 20 mM Tris−HCl, pH 7.5 and 20 mM NaCl at 54°C (4°C below T_m) to allow preferential hybridization of the oligonucleotide pairs containing single point mutations in complementary positions. Oligonucleotides were then ligated into *Sac*I−*Xho*I-restricted pU2.SP purified on agarose gel. Colony hybridizations were performed with a 'wild-type' oligonucleotide. After stringent washing with 6 × SSC at 56°C, the negative colonies, which corresponded to mutants, were picked and characterized by sequence analysis.

Some of the point mutants and the constructs pU2.SP-CCC and pU2.SP-USErev were generated by exchanging the *Sac*I−*Xho*I and, in some instances, the *Bam*HI−*Cla*I fragments of the synthetic promoter in pU2.SP for mutated sequences, using defined linker insertions. Complementary oligonucleotides were annealed as described for the preparation of the synthetic promoter and ligated into properly restricted pU2.SP. Mutant pU2.SPΔUSE was prepared from pU2.SP by digestion with *Cla*I and religation. In the mutant pU2.SP-USErev the sequence AAAGTCC-CACATCG (pos. −81 to −68) is in reverse orientation.

Mutants in the TATA box region

Mutants pU2.SP-TATA1 through −TATA3 were generated by inserting appropriate linkers into *Nco*I−*Acc*I-restricted pU2.SP. Plasmid pU2.SP used for construction of the TATA mutants had the *Sal*I site removed from the pTZ polylinker region; the *Sal*I sequence is recognized by *Acc*I nuclease. Construct pU2.SP-TATA4 was obtained by insertion of a 22 bp linker (Figure 5) into the *Acc*I site of pU2.SP.

Spacing mutants

Constructs pU2.SPace−1 through +2 were generated by inserting appropriate linkers into *Xho*I−*Nco*I-restricted pU2.SP. The pU2.SPace−2 was obtained by digestion of pU2.SP with *Xho*I and *Nco*I followed by filling in and religation. To obtain pU2.SPace+5, the 77 bp *Xho*I−*Dra*I fragment originating from the synthetic gene *syn7* (Goodall and Filipowicz, 1989) was cloned into pU2.SP pre-digested by *Nco*I and mung bean nuclease, and then by *Xho*I. Since blunting with mung bean nuclease resulted in removal of three additional nucleotides (GTA, positions −35 to −37; see Figure 6C), the spacing in pU2.SPace+5 was increased by 51 bp.

Transient expression in plant protoplasts

Transfection of plasmid DNA into protoplasts of *N.plumbaginifolia* was carried out using the polyethyleneglycol method as described previously (Negrutiu *et al.*, 1987; Vankan *et al.*, 1988). Unless indicated otherwise, 15 μg of plasmid DNA was used per transfection of 6×10^5 protoplasts. RNA was prepared 24 h after transfection according to the guanidinium−phenol−chlorophorm extraction method of Chomczynski and Sacchi (1987). RNA was treated with DNase I (RNase-free, from Promega Biotec) to remove any traces of remaining plasmid.

SP6 transcription and RNase A/T_1 mapping

RNA transcripts complementary to U2 RNA were synthesised *in vitro* by SP6 polymerase using [α-^{32}P]GTP (sp. act. 400 Ci/mmol) and *Hin*dIII-linearized pGEM2.U2.2 plasmid (Vankan and Filipowicz, 1988). A specific complementary probe used for mapping of U2 RNA transcribed from construct pU2.SP-TATA4 was obtained by the SP6 polymerase transcription of the U2.SP-TATA4 insert (*Bam*HI−*Eco*RI fragment) cloned into pGEM2.

RNase A/T_1 mapping was carried out as described previously (Vankan and Filipowicz, 1988), except that the RNase H treatment was omitted. Three micrograms of protoplast RNA were used for each assay and protected fragments were analysed on 6% polyacrylamide 8 M urea gels. Protected fragments corresponding in size to full-length U2 RNA were quantitated by scintillation counting. Values were normalized using the endogenous U2 RNA-protected fragments (80−90 nt long bands, see Figure 3A) as reference.

Acknowledgements

We thank J.Petruska for preparations of protoplasts, W.Zuercher and J.Jiricny for providing us with oligodeoxynucleotides and helpful discussions, F.Waibel for help with some experiments, and G.Goodall, F.Nagy and K.Wiebauer for critical reading of the manuscript.

References

Ares,M.,Jr, Mangin,M. and Weiner,A.M. (1985) *Mol. Cell. Biol.*, **5**, 1560−1570.

Ares,M.,Jr, Chung,J.-S., Giglio,L. and Weiner,A.M. (1987) *Genes Dev.*, **1**, 808−817.

Birnstiel,M.L. and Schaufele,F.J. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer Verlag, Berlin, pp. 183−195.

Bogenhagen,D.F. and Brown,D.D. (1981) *Cell*, **24**, 261−270.

Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156−159.

Ciliberto,G., Buckland,R., Cortese,R. and Philipson,L. (1985) *EMBO J.*, **4**, 1537−1543.

Ciliberto,G., Palla,F., Tebb,G., Mattaj,I.W. and Philipson,L. (1987) *Nucleic Acids Res.*, **15**, 2403−2416.

Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer Verlag, Berlin, pp. 38−70.

Dahlberg,J.E. and Schenborn,E.T. (1988) *Nucleic Acids Res.*, **16**, 5827−5840.

Goodall,G.J. and Filipowicz,W. (1989) *Cell*, **58**, 473−483.

Gram-Jensen,E., Hellung-Larsen,P. and Frederiksen,S. (1979) *Nucleic Acids Res.*, **6**, 321−330.

Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.*, **22**, 387−419.

Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562−8566.

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

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

Hinnebusch,A.G. and Fink,G.R. (1983) *J. Biol. Chem.*, **258**, 5238−5247.

Hutchison,C.A.,III, Nordeen,S.K., Vogt,K. and Edgell,M.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 710−714.

Janson,L., Bark,C. and Petterson,U. (1987) *Nucleic Acids Res.*, **15**, 4997−5016.

Joshi,C.P. (1987) *Nucleic Acids Res.*, **15**, 6643−6653.

Kazmaier,M., Tebb,G. and Mattaj,I.W. (1987) *EMBO J.*, **6**, 3071−3078.

Krol,A., Lund,E. and Dahlberg,J.E. (1985) *EMBO J.*, **4**, 1529−1535.

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

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

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mattaj,I.W. (1986) *Cell*, **46**, 905−911.

Mattaj,I.W., Lienhard,S., Jiricny,J. and De Robertis,E.M. (1985) *Nature*, **316**, 163−167.

Murphy,J.T., Skuzeski,J.T., Lund,E., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1987) *J. Biol. Chem.*, **262**, 1795−1803.

Mühlbach,H.P. and Sänger,H.L. (1979) *Nature*, **278**, 185−188.

Negrutiu,I., Shillito,R.D., Potrykus,I., Biasini,G. and Sala,F. (1987) *Plant Mol. Biol.*, **8**, 363−373.

Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259−266.

Parker,R., Simmons,T., Shuster,E.O., Siliciano,P.G. and Guthrie,C. (1988) *Mol. Cell. Biol.*, **8**, 3150−3159.

Parry,H.D., Scherly,D. and Mattaj,I.W. (1989a) *Trends Biochem. Sci.*, **14**, 15−19.

Parry,H.D., Tebb,G. and Mattaj,I.W. (1989b) *Nucleic Acids Res.*, **17**, 3633−3639.

Patterson,B. and Guthrie,C. (1987) *Cell*, **49**, 613−624.

Reddy,R. and Busch,H. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer Verlag, Berlin, pp. 1−37.

Skuzeski,J.M., Lund,E., Murphy,J.T., Steinberg,T.H., Burgess,R.R. and Dahlberg,J.E. (1984) *J. Biol. Chem.*, **259**, 8345−8352.

Southgate,C. and Busslinger,M. (1989) *EMBO J.*, **8**, 539−549.

Takahashi,K., Vigneron,M., Matthes,H., Wildeman,A., Zenk,M. and Chambon,P. (1986) *Nature*, **319**, 121−126.

Tanaka,M., Grossniklaus,U., Herr,W. and Hernandez,N. (1988) *Genes Dev.*, **2**, 1764−1778.

Vankan,P. and Filipowicz,W., (1988) *EMBO J.*, **7**, 791−799.

Vankan,P., Edoh,D. and Filipowicz,W. (1988) *Nucleic Acids Res.*, **16**, 10425−10439.

van Santen,V.L. and Spritz,R.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9094−9098.

van Santen,V.L., Swain,W. and Spritz,R.A. (1988) *Nucleic Acids Res.*, **16**, 41−76.

Wasylyk,B. (1988) *Biochim. Biophys. Acta*, **951**, 17−35.

Westin,G., Lund,E., Murphy,J.T., Petterson,U. and Dahlberg,J.E. (1984) *EMBO J.*, **3**, 3295−3301.

Wu,L. and Berk,A. (1988) *Genes Dev.*, **2**, 403−411.

Zoller,M. and Smith,M. (1983) *Methods Enzymol.*, **100**, 468−500.

Received on July 12, 1989