The EMBO Journal vol.7 no.3 pp.791 – 799, 1988

Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts

Pierre Vankan and Witold Filipowicz

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

Communicated by H.J.Gross

We have characterized the U2 snRNA gene family in the higher plant *Arabidopsis thaliana*. It consists of 10–15 genes which do not appear to be closely clustered. Six of the U2 genes were sequenced and the structure of the *Arabidopsis* U2 RNA termini was determined in order to define the coding regions. Each of the genes codes for a distinct RNA differing from the others by 2–13 point mutations, localized in the 3′ part of the 196 nt-long RNA. The upstream non-coding regions of all genes show strong sequence similarity in positions −81 to −1 and contain three highly conserved sequence elements: GTCCCACATCG (positions −78 to −68; 100% conservation), GTAGTATAAATA (−37 to −26) and CAANTC (−6 to −1). The coding regions are followed by the sequence $CAN_{7-9}AGTNNAA$, a putative termination signal. The expression of three of the genes was studied in electroporated *Orychophragmus violaceus* and *Nicotiana tabacum* protoplasts. The genes, one of which contains a T → C change in the Sm antigen binding site, were actively transcribed and processed into U2 RNAs of the expected size and containing trimethylguanosine caps. Deletion analysis indicates that sequences upstream of the conserved −80 to −1 region are not important for transcription in protoplasts. The 5′-terminal parts of U2 RNAs from several monocot and dicot plants were sequenced. This region, containing the sequence implicated in base-pairing with the branch point in pre-mRNA introns, is identical in all U2 RNAs examined.

Key words: Arabidopsis/plant gene expression/protoplast transfection/snRNA genes/U2 RNA

Introduction

All eucaryotic cells contain several abundant small nuclear RNA species known as U-snRNAs. These RNAs are organized in ribonucleoprotein particles (U-snRNPs) which are involved in various aspects of RNA processing in the nucleus (for review see Brunel *et al.*, 1985; Maniatis and Reed, 1987). In mammalian cells the U1, U2, U4, U5 and U6 snRNPs participate in pre-mRNA splicing as parts of the spliceosome (Maniatis and Reed, 1987). Experiments, carried out with mammalian and yeast extracts, indicate that U1, U5 and U2 snRNPs are involved in the recognition of the 5′ splice site, the 3′ splice site and the branch point region, respectively (Mount *et al.*, 1983; Black *et al.*, 1985; Chabot *et al.*, 1985; Parker *et al.*, 1987).

The synthesis of U-snRNAs in vertebrates has been extensively studied (reviewed by Dahlberg and Lund, 1987). All these RNAs, with the exception of U6, appear to be synthesized by RNA polymerase II; transcription of the U-snRNA genes is sensitive to α-amanitin (Gram-Jensen *et al.*, 1979) and the primary transcripts contain a 7-methylguanosine (m^7G) cap which is further modified in the cytoplasm to the 2,2,7-trimethylguanosine (m_3G) structure (Mattaj, 1986). However, the transcription signals of the U-snRNA genes differ from those found in protein coding genes. The U-RNA gene promoters contain two major sequence elments: the enhancer-like distal element positioned 200–250 nt upstream of the cap site and the −60/−50 proximal element responsible for accuracy of initiation (for recent refs see Kazmaier *et al.*, 1987; Murphy *et al.*, 1987; Dahlberg and Lund, 1987). This latter element is clearly distinct in its sequence and location from the TATA box which is responsible for selection of the start site in protein coding genes. Unlike most mRNAs, the U-snRNAs are not polyadenylated. The 3′ non-coding regions of their genes contain a 12–15 bp-long sequence which may be important for transcription termination or 3′ end processing (reviewed by Dahlberg and Lund, 1987). The differences in U-RNA and mRNA gene transcription are further underscored by the observation that the recognition of this 3′ end signal occurs only when transcription is initiated from an authentic U-RNA gene promoter and not from an mRNA promoter (Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986). These results suggest that snRNA genes are transcribed by a specialized transcription complex distinct from the RNA polymerase II complex which synthesizes mRNAs.

These peculiar features of U-snRNA gene transcription in animal cells prompted us to investigate the structure and activity of the U-snRNA genes in higher plants. Several RNAs belonging to the U1–U5 class have been previously identified in plants and some of them have been partially (U1, U2) or completely (U3, U5) sequenced (Krol *et al.*, 1983; Skuzeski and Jendrisak, 1985; Kiss *et al.*, 1985). However, the structure of the plant genes coding for U-RNAs have not been reported. We have focused our interest on the U2 RNA gene family in *A.thaliana* for the following additional reasons: (i) we have recently found that the mechanism of 3′ splice site selection in nuclear pre-mRNAs differs between plants and animals (K.Wiebauer, J.Herrero and W.Filipowicz, manuscript submitted). We are interested to find out whether this specificity is associated with the function of the U2 and U5 snRNPs, known to be involved in the recognition of the 3′ splice site region in introns of mammals or yeast (Black *et al.*, 1985; Chabot *et al.*, 1985; Parker *et al.*, 1987), and (ii) the extremely small genome of *Arabidopsis* (Leutwiller *et al.*, 1984) should facilitate thorough analysis of the gene family.

Results

Isolation of Arabidopsis U2 RNA genes

A genomic DNA library of *Arabidopsis* was screened with an SP6 polymerase transcript of the *Xenopus laevis* U2 RNA gene (Mattaj and Zeller, 1983) as the hybridization probe. Seven positive phages (U2.2−7 and U2.9) were plaque-purified and their DNAs analyzed by restriction enzyme digestion. Phage U2.6 appeared to be identical with the U2.3; each of the remaining six phages contained a single and distinct DNA fragment hybridizing to the probe. The fragments were subcloned into the pTZ18 vectors, mapped with restriction nucleases (Figure 1) and the shortest (0.8 kb) insert, originating from phage U2.2, was sequenced. The 0.8 kb fragment consists of a U2 RNA coding region (196 nt) flanked by 388 and 264 nt of upstream and downstream sequences, repectively (Figures 1 and 2). The coding region of the U2.2 gene contains a stretch of 62 nt (positions 7−68) identical to the sequence of *Xenopus* and other animal U2 RNAs (Reddy, 1986). Since it was likely that this region is also conserved in different *Arabidopsis* U2 genes, the genes U2.3, U2.4, U2.5, U2.7 and U2.9 and their flanking

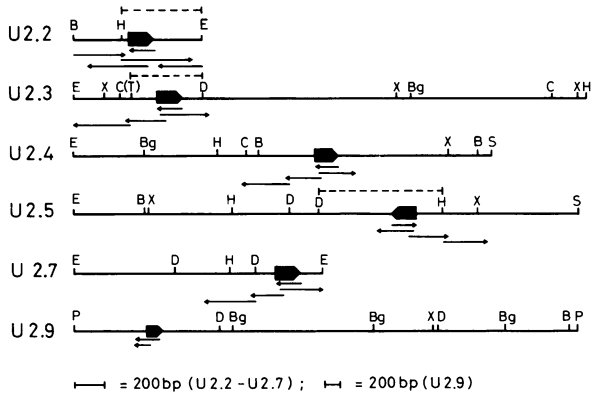

Fig. 1. Restriction maps of the 0.8−7.0 kb DNA fragments containing different U2 RNA genes. Thin arrows indicate sequenced regions. Arrow shaped boxes indicate coding regions and direction of transcription. Broken lines above U2.2, U2.3 and U2.5 indicate the DNA fragments that were transcribed to yield antisense RNA probes. B, Bg, C, D, E, H, P, S, T and X are *Bam*HI, *Bgl*II, *Cla*I, *Dra*I, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, *Taq*I and *Xba*I sites, respectively. The site in brackets is inferred from sequence analysis and is not unique.

```
          -190       -180       -170       -160       -150       -140       -130       -120       -110       -100
2.3 CACGAATCTGTCTACACCGATTCGATAGCTCCAGTCGAAGTTGTTTGTAGCTTCAAGGATTTGGTGAGAAGAAAACGTACGGATGGGATTGTATTTACCT
2.2 GATGCATTAAAAGACAAATTTATTCATTGTTTTGTATTGATAGTTCATATGTAATTGATTAAAGAAGAAATATGCGTAAGAGAAGACTAAAGAGTACATT
2.4 CAAAATCTCCGACAAGAACCAGATGCTAAGCCCATGACATATATTGAACCAGATGTCTAATGTCTCTTATTGATATGTTCGCGTTTGCTTATAGATGCGT
2.9 TATTAGAGCCGACAAGTACCAGATGCTAAGCCCATGACGAAAGTTGAAGCAGATGTCTGATGTCTTATTGATTTGTCGCGTTTGCTCATAGAGATGCGTA
2.7 AAATCTGATATTATCTGGGCTACGTGTGGATTAAAAATCCGATTAGGGGGAGTCGTGCAAAGCCCACTAAAAGAAACAGGCGGTTTGTTTTTATTTTAAA
2.5 GTTAAAAGAAGTTGGAAGCTTGTTTGGCTTGAGGTTTAAGACTATCATCAATCAGGCGTTGTTACGAAACCCGCTAAAAGAGAAAGAGGGTTATAGTTAT

           -90        -80        -70        -60        -50        -40        -30        -20        -10         -1
                 oo    ooo●●●●●●●●●●●●  ●●o  o  o●    o    ●  ●●     ●●ooo●●●●oo o●  oooo   ●  ●  ooo   o●●  ●●
2.3 CCTGACGATGCAATGAGTAAAGTCCCACATCGACAAGTTAGAGAGCCGTGGTCATAAGCATTGTAGTATAAA TAAGAACGTCCGCCCATCGTCCAAATC
2.2 TGGCTAAACCCTTTTTTAAAAGTCCCACATCGACAAGCTAGGGATCGTTGGTCATAAGCTTTGTAGTATAAA TAAGAAGCTGAGCCATTCGTTCAATTC
2.4 AAAGAGTCTATTTTTCTAAACGTCCCACATCGACAAGTTAGAGAGCATTGGTCATAAGATGTGTAATATATA AAGAAAGCTGAGCCAATTGTCCAATTC
2.9 AAAAATTCTAATTTTCTAAGAGTCCCACATCGACAAGCTAGAGAGAGTTTGTCATAACCTTAGTAGTATAA GTAAGAAGCTGAGCCCATCGAAAAATTC
2.7 TAACTTTTAACTCGTTCCAAAGTCCCACATCGCTAACAATTTAAGAGGTGAAGACAAGACGAGTAGTATAAA TAACGAGCCACGGCCA CGAACAATTC
2.5 TTAGTTTCACTGGTTCACGAAGTCCCACATCGCTAAGAAA AGAACGTGAAAGAAAAAACGAGTGGTATAAAGTACGAGGCCATGGCACTC AGCAAATC

  * 1        10         20         30         40         50         60         70         80 *       90         100
All ATACCTTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAATATCTGATATGTGGGCCATCGGCCCACACGATATTAACTCTA

          110        120        130        140        150        160        170        180        190
2.3 TTTTTTAAGGGAGAAAGCCCACTAAGATAGCTTGCTATCTGGGCTTTCACGAGTCGCCCATGCGTTGCACTACTGCACGGGCCTGGCTCAACCCGC
2.2 .C.................G...T.....................C.....................................................
2.4 ..................................................A.....................................-.............
2.9 ....................GT.........................G...........
2.7 ................A.........G................T.....................................................T.....
2.5 ............G......GT.T.....................................T................T...GA....T.............

     200        210        220        230        240        250        260        270        280        290
     ●●   oo    ●●● o●●   o  o o
2.3 CAAGCAAATAAAGTCAAAACTTTTTAAACCTTACTCTTTATCTTTCAAGTCCAAAAGCTCAGAGGATAAAAGTGAATGGAGAAAACTTCGTTAAATTATC
2.2 CAAACTAGA  AGTGAAACTTTTAAAATTTTAGTCACTTTTGATGTTTAGTTTTTAAAAATGATGTAGGATTCTTTATCATTCAAAGTCCAAAATATACG
2.4 CAGCTAACC  AGTCAAATTTTCGACTCATCTCTAATTAATCAAACTCTATAAAACCAAAACTTAATTAATTAACACCAAATCAACAATGAACAGAGGAA
2.7 CAATAAACC  AGTAAAATTATAAATATTGGAAATCTGCCACTATGTTGTAGATTTGGGCTTAGGGACCTAAATAAATAACGCCATGTAACTGGTGGATG
2.5 CATTTATTC  AGTCTAAGGAAATAAAAAAAAGCTAAAGTGGATTAAATCGAAAAGCCAAATAAAATCCATGAACGTCCGTCTTTCTTGAAGCCCATGAC
```

Fig. 2. Sequences of U2 RNA genes of *Arabidopsis*. Numbering corresponds to the gene U2.3. Other sequences are aligned with U2.3 allowing some gaps in order to accentuate similarities in the noncoding regions. Dots indicate identical positions in the coding region. (−) in the gene U2.4 indicates a deletion. (*) a T residue is present in position 82 in the U2.2 gene. Closed and open circles indicate identity in six or five out of the six sequences in the promoter region and identity in five or four out of the five sequences in the downstream region. The far-upstream sequences conserved between genes U2.2 and U2.3, U2.4 and U2.9, and U2.7 and U2.5 are indicated by dotted, continuous and broken lines, respectively. Only part of the sequenced noncoding DNA is shown. The complete upstream and downstream sequences (U2.2 gene, 388 and 264; U2.3, 511 and 160; U2.4, 588 and 101; U2.5, 455 and 97; U2.7, 502 and 138; U2.9, 199 nt) are available upon request and will be submitted to the EMBL Data Bank. The 3′ terminal part of the U2.9 gene has not been sequenced.

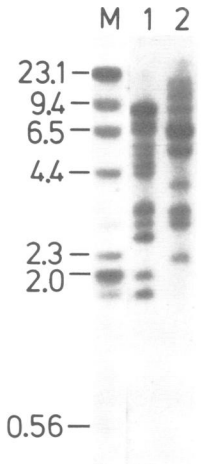

Fig. 3. Southern blot analysis of *Arabidopsis* DNA, restricted with *PstI* (**lane 1**) and *PvuII* (**lane 2**). DNA was probed with an antisense transcript of the cloned U2.2 gene.

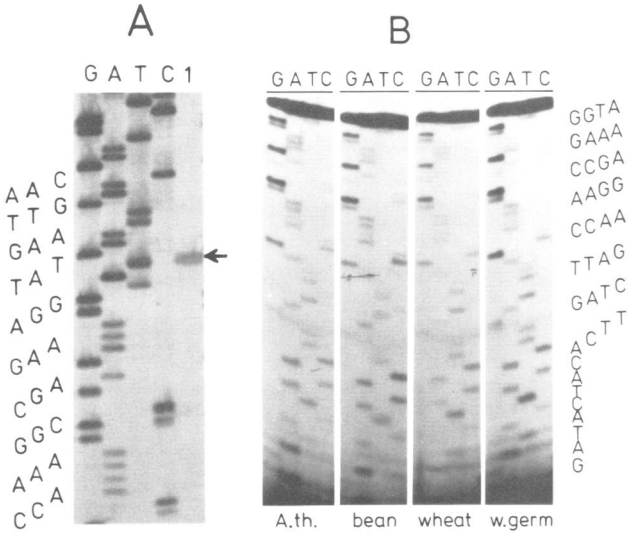

Fig. 4. (**A**) Analysis of the *Arabidopsis* U2 RNA 5′ end by primer extension (**lane 1**). The product of reverse transcription is indicated by an arrow. **Lanes G, A, T** and **C** show the sequencing reactions of the pTZ18U.U2.2 ssDNA. (**B**) Sequences of the 5′ termini of U2 RNAs isolated from *Arabidopsis thaliana*, broad bean, wheat seedlings (wheat) and wheat germ. The sequences complementary to U2 RNA are indicated.

regions were sequenced starting with internal oligonucleotide primers complementary to this region (Figures 1 and 2).

Southern blot analysis of *Arabidopsis* DNA digested with different restriction enzymes indicated the presence of 10–12 bands hybridizing to an RNA probe originating from the U2.2 gene (Figure 3). Since each of the phages characterized above contains only a single U2 gene, it is likely that the U2 RNA genes of *Arabidopsis* are not closely clustered and that their total number corresponds to the number of hybridizing fragments seen in Southern analysis. This estimation is supported by the number of positive plaques seen during library screening. The results do not, however, eliminate the possibility that one or more of the non-characterized DNA fragments contain two or three genes positioned closely together or that some of the bands seen in Southern analysis are due to allelic polymorphism.

Structure of the coding region

In order to define the coding region of the U2 genes, the 5′ and 3′ termini of the *Arabidopsis* U2 RNAs were analyzed. An RNA preparation enriched in U2 RNA was obtained by immunoprecipitation with anti-m_3G cap antibody (Lührmann *et al.*, 1982). The 5′ end of U2 RNA was determined by reverse transcription in the presence (sequencing reactions; Figure 4B) or absence (primer extension, Figure 4A) of dideoxynucleotides, using a 16 nt primer complementary to positions 47–62 of *Arabidopsis* U2 RNA. The results indicated that the first transcribed nucleotide corresponds to the A at position 1 (Figure 2). Based on these data and the immunoprecipitability of the RNA with anti-m_3G antibody, it is concluded that the 5′ end corresponds to $m_3GpppAU$-ACC. To determine the nature of the 3′ end, *pCp-labeled, hybrid-selected U2 RNA of *Arabidopsis* was digested with RNase A, T1 or T2, and the products analyzed by TLC in different systems (see Materials and methods). The results are consistent with ~50% of the U2 RNA molecules having a terminus $CCCGC_{OH}$ ($C\overset{*}{p}Cp$, $GpC\overset{*}{p}$ and $C\overset{*}{p}$ produced after digestion with RNase T1, A and T2, respectively; data not shown), similar to the termini of U2 RNAs isolated from pea and broad bean ($CCCAC_{OH}$; Krol *et al.*, 1983; Kiss *et al.*, 1985) or wheat ($CCCAA_{OH}$; Skuzeski and Jendrisak, 1985). The 3′ ends of the other 50% of the *Arabidopsis* U2 RNA population were not precisely assigned The presence of C residue at the 3′ end and generation of long (7–8 nt) oligonucleotides by digestion with RNase T1 are suggestive of termination at the C residue(s) positioned upstream of G_{195} (data not shown and Figure 2).

The first 82 nt of the coding region are identical in all sequenced U2 RNA genes but several nucleotide substitutions are present further downstream. For example, genes U2.5 and U2.7 differ at 13 positions, whereas genes U2.3 and U2.4 only differ at two positions. The U2.2 gene contains a $T_{102} \rightarrow C$ change in the sequence corresponding to the Sm antigen binding domain (AUUUUUUA) present in U2 and other U-RNAs (Branlant *et al.*, 1982; Reddy, 1986).

The RNAs encoded by *Arabidopsis* U2 genes can be folded into the secondary structure model proposed for animal U2 RNAs (Keller and Noon, 1985) (Figure 5). With the exception of a $U_{147} \rightarrow C$ substitution in the RNA encoded by the U2.2 gene, the nucleotide changes would not disturb the secondary structure of the U2 RNA. Changes in helical stems preserve the base pairing, while others occur either in non base-paired regions or are immediately adjacent to single-stranded bulges in stems ($A_{127} \rightarrow G$ in U2.7 and $G_{156} \rightarrow U$ in U2.5).

Structure of the 5′ and 3′ non-coding regions

The upstream non-coding regions in all six genes show significant sequence similarity in positions −1 to −81 (Figure 2). Based on the degree of sequence conservation in this region the genes can be subdivided into two groups: one includes genes U2.2, U2.3, U2.4 and U2.9 (74–83% similarity) and the other genes U2.5 and U2.7 (69% similari-

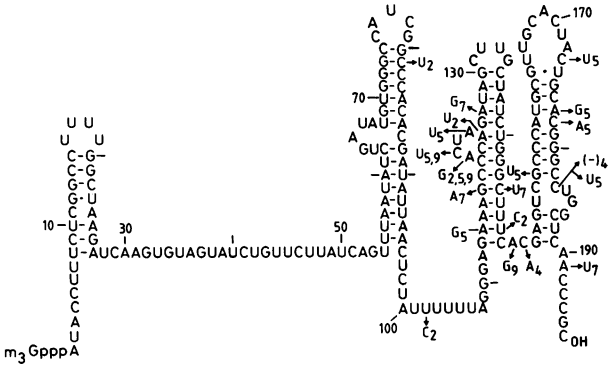

Fig. 5. Secondary structure of *Arabidopsis* U2 RNAs deduced from DNA sequences. The structure of the U2.3 RNA is drawn according to the model of Keller and Noon (1985). The nucleotide changes inferred from the sequences of other cloned U2 RNA genes are indicated. About half of the *Arabidopsis* U2 RNA population may be 2–3 nt shorter at the 3′ end (see Results).

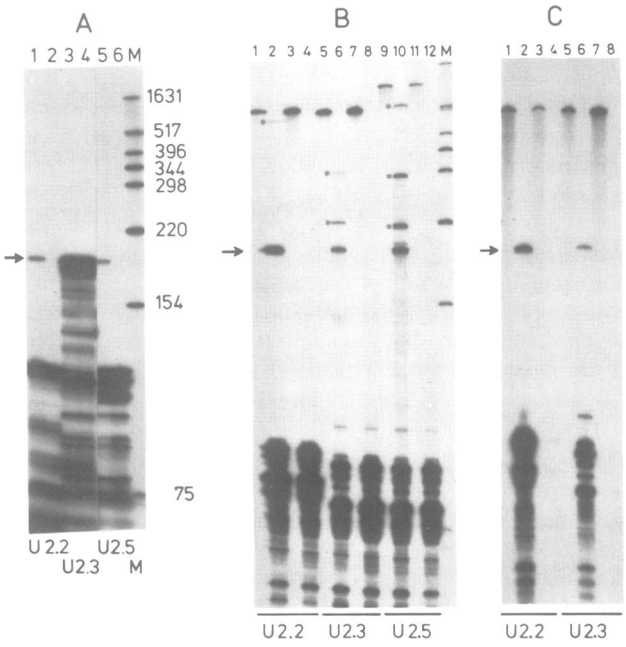

Fig. 6. Expression of the *Arabidopsis* U2 RNA genes in transfected *Orychophragmus* and tobacco protoplasts. (**A**) **Lanes 1, 3** and **5**, mapping of RNA from *Orychophragmus* protoplasts transfected with plasmids pTZ18U.U2.2, pTZ18U.U2.3 and pTZ18U.U2.5, respectively. **Lanes 2, 4** and **6**, analysis of RNAs from mock transfected protoplasts. (**B**) **Lanes 2, 6** and **10**, mapping of RNAs from tobacco protoplasts transfected with plasmids pTZ18U.U2.2, -U2.3 and -U2.5, respectively. **Lanes 4, 8** and **12**, analysis of RNAs from mock transfected protoplasts. **Lanes 1, 3, 5, 7, 9** and **11**, aliquots of undigested probes used in each analysis. (**C**) Analysis of RNAs immunoprecipitated with anti-m_3G (**lanes 2** and **6**) or control (**lanes 4** and **8**) antibody prior to RNase A/T1 mapping. **Lanes 1, 3, 5** and **7**, aliquots of undigested probes, used in each analysis. The gene-specific RNA probes used are indicated at the bottom and schematically shown in Figure 1 (broken lines). Arrows indicate the position of mature U2 RNA and asterisks in panel C indicate the longer protected RNAs. **Lanes M**, size markers (3′ end-labeled *Hin*fI digest of pBR322).

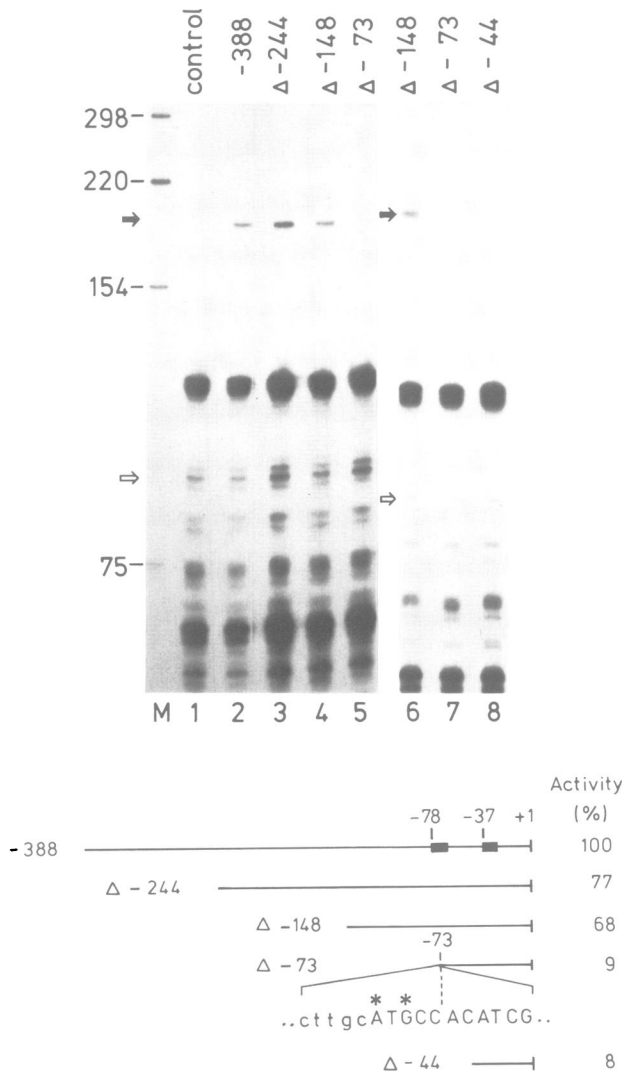

Fig. 7. Activity of the U2.2 gene with promoter deletions. Plasmids containing the U2.2 RNA gene with progressive 5′ deletions (schematically shown below the gel and identified above each lane) were transfected into *Orychophragmus* protoplasts. Filled arrows, position of mature U2 RNA; open arrows, 'endogeneous' protected RNA band used for normalization of the U2.2 gene expression. **Lane M**, size markers; **lane 1**, analysis of RNA from mock transfected protoplasts. **Lanes 1–5** and **6–8** originate from separate experiments. The positions of the two conserved upstream boxes are indicated in the diagram. In the construct Δ-73 the plasmid sequence ligated to the site of deletion (position −73) restores the original sequence of the box, apart from two positions (marked with asterisk). The activity of each mutant, expressed as a percentage of the activity of the wildtype plasmid (−388, corresponding to pTZ18U.U2.2), is an average from two independent transfections. The intensity of bands was quantitated by densitometry.

ty). The sequence similarity between genes belonging to different groups is 56–61%. The region −81 to −1 contains three sequence elements that are conserved in all the U2 RNA genes: a GTCCCACATCG box (100% conservation, positions −78 to −68); a TATA motif-containing sequence GTAGTATAAATA (−37 to −26) and the cap site-adjacent sequence CAANTC (Figure 2); these sequences may be of importance for initiation of transcription (see below). Several other sequences common to the genes belonging to one subfamily are also present in this region. They may reflect the evolutionary history of the genes.

Further upstream the sequence similarity between different U2 RNA genes is much less pronounced. Sequences common to all the genes were not identified in the regions which

have been sequenced (199 bp in the U2.9 gene and at least 388 bp in all the others, see legend to Figure 2). However, in the −80 to −190 region several identical or similar sequences are shared by pairs of genes: U2.2 and U2.3, U2.4 and U2.9, and U2.5 and U2.7 (Figure 2). The upstream homology between U2.4 and U2.9 is particularly extensive. It remains to be established whether sequence similarities between different U2 RNA genes are important for their activity or just reflect the common origin of these genes and a recent duplication event in the case of the U2.4/U2.9 pair.

The similarity in the 3′ non-coding regions is restricted to the sequence $CAN_{7-9}AGTNNAA$, immediately following the end of the coding sequence in all genes (Figure 2). This sequence may play a function during 3′ end formation.

Expression of the U2 RNA genes in transfected protoplasts

The activity of genes U2.2, U2.3 and U2.5 was tested by measuring their transient expression in *Orychophragmus violaceus* [a plant belonging to the same (*Cruciferae*) family as *Arabidopsis*] and tobacco protoplasts. Total protoplast RNA was isolated 24−35 h after transfection and analyzed by RNase A/T1 protection, using ^{32}P-labeled antisense RNA probes specific for each gene. It was assumed that the U2 RNAs of *Orychophragmus* and tobacco would differ from those encoded by *Arabidopsis* genes and thus would not form perfect RNase-resistant hybrids with the probes. This assumption proved to be valid for two out of three genes transfected into *Orychophragmus* protoplasts and for all three genes transfected into tobacco protoplasts. RNA isolated from *Orychophragmus* protoplasts transfected with the U2.2 and U2.5 genes (but not from control protoplasts) protected RNA fragments corresponding precisely in length to mature U2 RNA (Figure 6A, lanes 1 and 5). The U2.3 gene expression could not be measured because the *Orychophragmus* cells contained an endogenous U2 RNA species which fully protected the U2.3 probe (lane 4). In tobacco protoplasts, after transfection with each of the genes, a new protected RNA band corresponding in size to mature U2 RNA was observed (Figure 6B, lanes 2, 6 and 10). Protected RNA fragments longer than the U2 RNA were also formed. The U2.2 gene-specific ~450 nt fragment (lane 2) is most probably diagnostic of an RNA initiated at position +1 but extending beyond the end of the RNA probe used for RNase mapping (the U2.2-specific probe is 503 nt long with 42 nt complementary to the upstream non-coding region, see Figure 1). Additional fragments observed after transfection with the genes U2.3 and U2.5 are also likely to correspond to RNA transcripts which are not properly terminated and/or processed in tobacco cells, although other explanations of their origin are possible. Mapping with additional probes is required to settle this point. It should be noted that fragments longer than mature U2 RNA were never observed when RNA from transfected *Orychophragmus* protoplasts was analyzed (Figures 6 and 7, and data not shown), indicating that some of the signals for U2 RNA synthesis may differ between different plant families.

To determine whether the U2 RNA gene transcripts contain mature 5′ ends capped with m_3G, the RNA isolated from tobacco protoplasts transfected with the U2.2 and U2.3 genes was precipitated with anti-m_3G antibody prior to RNAse mapping. As seen in Figure 6C, the immunoselected RNAs protected the U2 RNA-length fragment but not any of the longer fragments discussed above. RNAs treated with control antibody did not yield any protected fragments.

Having established that transcription of exogenous U2 genes into products with properties of mature U2 RNA can be measured in protoplasts, we studied the effect of deletions in the 5′ flanking region on the activity of the U2.2 gene. The U2.2 gene was selected because out of the three genes tested it was the most efficiently transcribed (Figure 6). Comparison of the activity of the original U2.2 construct (containing 388 nt of the upstream sequence) with the mutants containing progressive deletions in the upstream region, is shown in Figure 7. Deletions of 144 or 240 nt (mutants Δ-244 and Δ-148, respectively) had only negligible effect on gene activity. However, mutants Δ-73 (deletion reaching into the −78/−68 box), resulting in a change of two bases in the 11 nt sequence conserved in all U2 genes, and Δ-44 (deletion extending beyond the −78/−68 box) were about 10-fold less active than control plasmid. Residual transcription of these mutants still resulted in formation of the mature-sized U2 RNA (Figure 7 and data not shown). The results of deletion analysis suggest that the conserved −80 to −1 region contains all the elements required for efficient initiation of transcription in transfected protoplasts. Although the function of the −148 to −80 region is not directly tested in our experiments, it is unlikely that these sequences are of importance since this region does not contain any sequence elements conserved between more than two U2 RNA genes (Figure 2).

The 5′-terminal regions of monocot and dicot U2 RNAs

It has been demonstrated in yeast that the U2-like snRNA and the branch point intron sequence UACUAAC interact by base-pairing during pre-mRNA splicing (Parker *et al.*, 1987); a similar interaction between pre-RNA and the U2 RNA may also occur in higher eucaryotes (Keller and Noon, 1984; Black *et al.*, 1985; Parker *et al.*, 1987). The U2 RNA region that is implicated in base-pairing with pre-mRNA (it corresponds to nt 34−39 in *Arabidopsis* U2 RNA; see Parker *et al.*, 1987) is conserved between yeast (Ares, 1986), animals (Reddy, 1986) and a higher plant, *Arabidopsis*. However, partial sequencing of the three U2 RNA variants in wheat germ (Skuzeski and Jendrisak, 1985) has indicated that these RNAs contain two U → C changes (U_{35} and U_{38} in *Arabidopsis* RNA, corresponding to C_{30} and C_{33} in wheat RNA; the numbering in wheat RNA differs since some of the modified nucleotides have not been sequenced), exactly in the region implicated in interaction with pre-mRNA. We have reinvestigated the structure of wheat U2 RNA (isolated from germ and seedlings) and also analyzed the sequence of U2 RNAs isolated from other monocots (rice and barley seedlings, maize leaves and kernels) and dicots (broad bean). It was found that in all these RNAs the sequence of the forty 5′-terminal nucleotides is identical to the sequence of *Arabidopsis* U2 RNA (Figure 4B and data not shown).

Discussion

The results presented in this work indicate that *A.thaliana*, like other higher eucaryotes (reviewed by Dahlberg and

Lund, 1987), contains multiple U2 snRNA genes. However, in contrast to most of the vertebrates, *Drosophila* and sea urchin, in which the U2 genes are either clustered or organized in tandem repeats (Dahlberg and Lund, 1987), the *Arabidopsis* genes appear to be dispersed and are flanked by distinct upstream and downstream sequences; the U2 RNA genes in rodents may have a similar organization (Tani *et al.*, 1983). Each of the six cloned genes of *Arabidopsis* codes for a variant U2 RNA species. The following observations argue strongly in favor of the isolated genes being functional and not pseudogenes: (i) all six genes have conserved coding regions. The nucleotide changes in the RNA encoded by each gene maintain the secondary structure of U2 RNA; one of the genes (U2.7) contains two compensatory substitutions ($G_{117} \rightarrow A$ and $C_{144} \rightarrow T$). The $U_{147} \rightarrow C$ change in U2.2 RNA is the only exception to this rule, but efficient transcription of the U2.2 gene in protoplasts has been demonstrated; (ii) the upstream and downstream non-coding regions contain several elements which have conserved nucleotide sequences and positions within the genes. These sequences most likely represent transcriptional signals (see below); (iii) of the three genes used in protoplast transfection experiments, all three were actively transcribed into U2 RNA-sized products containing the m_3G cap; and (iv) RNase A/T1 analyses carried out with the RNA isolated from *Arabidopsis* indicated that U2 RNAs able to protect the U2.2, U2.3 and U2.5 antisense probes (only these probes were tested) are expressed in *Arabidopsis* (unpublished results).

If all of the described genes are indeed functional, the complexity of the *Arabidopsis* U2 RNA population would represent the most extreme case of U2 RNA heterogeneity known to date. The heterogeneity of U2 RNA may be a more general phenomenon, particularly in plants, since three U2 RNAs species have been identified in wheat (Skuzeski and Jendrisak, 1985). Four U2 genes present in *Drosophila* genome can potentially encode three RNA variants (Alonso *et al.*, 1984). On the other hand, only a single U2 RNA species has been so far detected in vertebrates (for refs see Reddy, 1986). Since the genome of *Arabidopsis* is extremely small (7×10^7 bp; Leutwiler *et al.*, 1984) and the multigene families are usually much smaller than in other plants (Leutwiler *et al.*, 1986 and refs therein) the large number of the genes encoding different U2 RNA species in *Arabidopsis* is surprising and raises the question of the significance of U2 RNA heterogeneity. It is possible that variant U2 RNAs are required for splicing of different introns or for alternative pre-mRNA processing in different tissues. However, if the two-dimensional structure rather than the nucleotide sequence of the 3′ half of U2 RNA is important for U2 snRNP assembly and function, the reported variability may have no importance. It should be noted that multiple variants of other U-RNAs have been described in different organisms (Krol *et al.*, 1983; Forbes *et al.*, 1984; Strub *et al.*, 1984; Lund *et al.*, 1985; Reddy, 1986) and that expression of different types of U1 and U4 RNAs in mice and frogs is developmentally controlled (Lund *et al.*, 1985; Lund and Dahlberg, 1987). The significance of these findings is unknown.

The *Arabidopsis* U2.2 gene contains a $T_{102} \rightarrow C$ substitution which would result in a change of the Sm antigen binding domain (AUUUUUUAG to AUCUUUUAG) present in U-RNAs (Branlant *et al.*, 1982). All U-RNA genes or U-RNAs studied so far contain a U residue in this position (Reddy, 1986). It was unexpected to find that transcription of the U2.2 gene in protoplasts results in accumulation of high levels of a U2-like RNA product bearing the m_3G cap. In *Xenopus* oocytes the hypermethylation of the m^7G cap occurs in the cytoplasm and depends upon prior binding of Sm protein(s) to RNA (Mattaj, 1986). If the maturation of U-RNAs follows the same pathway in plants, the results would indicate that the nucleotide sequence requirements for Sm antigen binding in plants are less stringent than in other organisms.

Prompted by the finding that the 5′ region of wheat germ U2 RNA (Skuzeski and Jendrisak, 1985) does not exactly conform to the sequence that is absolutely conserved in all other higher eucaryotes (including *Arabidopsis*) examined, and the implications this could have for the observed differences in pre-mRNA splicing between monocots and dicots (Keith and Chua, 1986), we have sequenced the 5′ termini of U2 RNAs from several monocot and dicot plants. It was found that the forty 5′-terminal nucleotides are conserved in all sequenced plant U2 RNAs, including those isolated from wheat seedlings and wheat germ. Hence, the region in U2 RNA which, based on experiments in the yeast *Saccharomyces cerevisiae* (Parker *et al.*, 1987), may interact with the branch-point sequence during pre-mRNA splicing, is conserved among all higher eukaryotes and yeast (Reddy, 1986; Ares, 1986). Other factors are therefore likely to be responsible for observed specificities of plant versus animal (Barta *et al.*, 1986; K.Wiebauer, J.Herrero and W.Filipowicz, manuscript submitted) and monocot versus dicot (Keith and Chua, 1986) pre-mRNA splicing.

Since this work was completed a note reporting the primary structure of U2 RNA from broad bean has appeared (Kiss *et al.*, 1987). The sequence of the 5′ end of broad bean U2 RNA determined by us (Figure 4B) differs from the reported sequence by the addition of a U residue at position 7. A comparison of bean U2 RNA sequence (which is likely to represent the 'consensus' sequence of a mixture of RNAs if this plant contains several U2 RNA species) with the RNA encoded by *Arabidopsis* genes indicates that it is most closely related to the U2.3 variant (84% similarity; 30 nt substitutions, 1 nt insertion, 1 nt deletion). Comparison of *Arabidopsis* U2 RNAs with vertebrate RNAs (Reddy, 1986) indicates about 68% similarity. *Arabidopsis* U2 RNA, like the U2 RNAs from other higher eucaryotes, can be folded into the secondary structure proposed by Keller and Noon (1985), which emphasizes the importance of this structure for U2 RNA function. The seventy 5′-terminal nucleotides represent the most conserved region in U2 RNAs from different organisms. Additional highly conserved sequences correspond to the Sm domain and to the single-stranded loops in two 3′-proximal hairpins of the RNA (positions 130−135 and 165−172 in *Arabidopsis* U2 RNA; see also Reddy, 1986 and Kiss *et al.*, 1987).

Comparison of upstream and downstream non-coding regions allowed us to identify sequence elements conserved in all *Arabidopsis* U2 genes. The upstream elements include: GTCCCACATCG (positions −78/−68, 100% conservation), GTAGTATAAATA (−37/−26) and CAANTC (−6/−1). Immediately adjacent to the 3′ end of the coding region is the sequence $CAN_{7-10}AGTNNAA$. Since the *Arabidopsis* U5 RNA gene also contains all the upstream and downstream elements specified above (D.Edoh and

P.Vankan, unpublished results), these sequences most probably represent signals required for U-RNA synthesis in plants. Interestingly, the spacing between the three upstream elements is conserved in all *Arabidopsis* genes. Approximately three helical turns separate the boxes −78/−68 and −37/−26 and two turns the boxes −37/−26 and −6/−1, suggesting that all these sequences may interact with components of transcriptional machinery on one side of the helix. The absence of common sequences upstream of −80, and the results of the 5′ deletion analysis of the gene U2.2, indicate that the −80/−1 region contains all the signals required for efficient initiation of transcription in protoplasts. However, it cannot be excluded at present that additional sequences, possibly positioned further upstream, are also of importance for transcription in plant tissues.

Analysis of the promoters in the vertebrate genes coding for U1, U2 and U5 RNAs indicated the presence of two major transcription signals: the enhancer-like distal element positioned 250−200 nt upstream from the cap site, and the −60/−50 proximal element which is responsible for correct initiation of transcription (reviewed by Dahlberg and Lund, 1987). A characteristic feature of all animal U-RNA genes is the lack of TATA box around position −30, which is important for the accurate initiation of mRNA synthesis. The conserved upstream sequence elements in *Arabidopsis* genes are distinct from their animal counterparts, with respect both to sequence and location. The upstream −78/−68 element may have an enhancing effect on transcription since modification or deletion of this element in the U2.2 gene resulted in ~10-fold decrease in U2 RNA synthesis. The sequence of the −78/−68 box also does not resemble any of the known promoter elements in plant mRNA genes (our unpublished observations). The most interesting feature of the *Arabidopsis* U-RNA genes promoter is the presence of the TATA homology-like sequence GTAGTATAAATA in the −30 region. Although this sequence is distinct from the TATA box consensus derived for protein-coding genes in plants (tcacTATATATAg; Joshi, 1987), it is likely that both elements are functionally related. The residual transcription of the 5′ deletion mutants in which the −78/−68 element is modified or deleted but the −37/−26 box remains intact, still resulted in formation of U2 RNA products of correct size. It will be of interest to find out whether the −30 regions in U-RNA and mRNA genes in plants are interchangeable.

We are presently carrying out a more detailed functional analysis of the conserved sequence elements in *Arabidopsis* genes, using the protoplasts of *Orychophragmus* and tobacco which represent *Cruciferae* and *Solanaceae* families, respectively. The observed differences in the transcription patterns of the *Arabidopsis* genes in these 'host' and 'non-host' systems (Figure 6) should help to elucidate the mechanism of U-RNA gene transcription and RNA maturation in plants.

Materials and methods

Materials

Cellulase R-10 and macerozyme R-10 were bought from Yakult Honsha Co. (Takarazuka, Japan). Pectolyase Y-23 and drisellase were from Seishin Pharmaceutical Co. (Noda, Japan) and Fluka, respectively. Radioisotopes were brought from Amersham. Oligodeoxynucleotides, synthesized on an Applied Boisystems synthesizer, were kindly provided by W.Zürcher and J.Jiricny of this Institute.

Construction and screening of the genomic library

Genomic DNA was prepared from 4-week-old *Arabidopsis* plants (Benscheim Be-O strain), obtained from I.Negrutiu (Free University of Brussels), using the method of Murray and Thompson (1980). An *Arabidopsis* genomic DNA library was constructed by ligation of partially *Sau*3A-digested, size-fractionated and dephosphorylated DNA with the arms of the λEMBL-3 vector (Frischauf *et al.*, 1983). The library was screened according to Benton and Davis (1977), using as a probe an antisense RNA transcript of the *Xenopus* U2 gene (Mattaj and Zeller, 1983). The plasmid pSP62AU2, used as a template, was obtained from I.Mattaj (EMBL, Heidelberg). A total of 18 000 plaques, corresponding to five *Arabidopsis* genome equivalents, were screened. Hybridizations were performed in 5 × SSC, 5 × Denhardt, 50 mM Na_2HPO_4, 250 μg/ml tRNA, 0.1% SDS for 16 h at 42°C. Filters were washed twice in 2 × SSC, 0.1% SDS at 50°C for 30 min. After Southern blot hybridizations filters were additionally washed in 0.1 × SSC, 0.1% SDS for 30 min at 68°C, before autoradiography.

Subcloning and sequencing strategy

Unless stated otherwise, all techniques for manipulating DNA were as described in Maniatis *et al.* (1982). The 0.8 kb *Eco*RI−*Bam*HI fragment from phage U2.2, which hybridized to the *Xenopus* probe, was isolated from the agarose gel and subcloned into the plasmids pTZ18U and pTZ18R (Mead *et al.*, 1986). ssDNA was produced according to Dente *et al.* (1983), using M13K07 as the superinfecting phage. The sequence was determined by the dideoxynucleotide method using [α-^{35}S]ATP and either the M13 universal primer or internal primers. Larger U2 gene-containing fragments from the phages U2.3 (*Eco*RI−*Hin*dIII, 3.5 kb), U2.4 (*Eco*RI−*Sal*I, 2.9 kb), U2.5 (*Eco*RI−*Sal*I, 3.5 kb), U2.7 (*Eco*RI, 1.7 kb) and U2.9 (*Pst*I, 7 kb) were subcloned into pTZ18U or pTZ18R, using shotgun cloning. Positive clones were identified by hybridization to the U2.2 gene probe. These clones were sequenced using the internal primers. Each region was sequenced at least twice.

The constructs used for SP6 transcription were prepared as follows: the 0.8 kb U2.2 insert was cloned into the *Bam*HI−*Eco*RI sites of pGEM2, to yield pGEM2.U2.2; a 0.5 kb *Dra*I−*Taq*I U2.3 fragment was first cloned into the *Acc*I−*Sma*I-cleaved Bluescript$^+$ vector (Stratagene) and then recloned after cleaving with *Bam*HI and *Xho*I, into *Bam*HI−*Sal*I-cleaved pGEM2 to yield pGEM2.U2.3; a 1 kb *Dra*I−*Hin*dIII U2.5 fragment was cloned into *Sma*I−*Hin*dIII-cleaved pGEM2 to yield pGEM2.U2.5. All plasmids were linearized with *Hin*dIII; in pGEM2.U2.3 this site is present in a polylinker. The regions transcribed into RNA are indicated in Figure 1.

Deletion mutants in the U2.2 gene were produced using the exonuclease III approach (Henikoff, 1984). The enzymes and the protocol followed were obtained from Stratagene. DNA from plasmid pTZ18U.U2.2 was digested with *Bam*HI and *Pst*I to generate the required ends. The deletion mutants were characterized by sequence analysis.

Preparation of RNA probes

Linearized DNA templates were transcribed *in vitro* with SP6 polymerase (Melton *et al.*, 1984), using [α-^{32}P]GTP (sp. act. 800 Ci/mmol). Transcripts used for RNase A/T1 protection assays were purified by gel electrophoresis.

Preparation of snRNAs from plants

Total RNA was prepared from 7-day-old seedlings of rice (var. M302), barley (var. Golden Promise), and wheat (var. Farnese), from 4-week-old *Arabidopsis* plants, from the leaves of 12-week-old maize plants (var. B73), from maize kernels and from commercial wheat germ (General Mills, Vellejo, California) by the method of Short and Torrey (1972). Small RNA was extracted from the ethanol pellets with 3 M Na acetate containing 5 mM EDTA, precipitated with ethanol and dissolved in H_2O. This RNA was used for immunoprecipitation with anti-m_3G antibodies (supplied by R.Lührmann, Max-Planck-Institute of Molecular Genetics, Berlin) according to Krol *et al.* (1983). RNA from nuclei of broad bean (*Vicia faba L.*) was kindly provided by T.Kiss and F.Solymosy, Institute of Plant Physiology, Szeged, Hungary.

Determination of the 3′ ends of Arabidopsis U2 RNA

Immunoprecipitated snRNA of *Arabidopsis* was labeled with [5′-^{32}P]pCp (sp. act. 3000 Ci/mmol) using T4 RNA ligase (Konarska *et al.*, 1981), and fractionated on a 10% polyacrylamide gel. The band corresponding to U2 RNA was eluted and RNA was further purified by hybrid selection as described by Murphy *et al.*, (1982) using the plasmid pTZ18U.U2.2. RNA was digested with RNase A, T1 or T2 and digests were analyzed by TLC on

cellulose plates in solvent A [saturated $(NH_4)_2SO_4$/1M Na acetate/isopropanol, 80:12:12] or B [isobutyric acid/NH_4OH/H_2O, pH 4.3, 577:38:385], or on DEAE–cellulose plates in solvent C [0.2 M NH_4 formate, 9 M urea, 1 mM Na_2 EDTA] after a short prerun in water (Konarska *et al.*, 1981).

Sequencing of the 5′ ends of U2 RNAs

Sequences at the 5′ ends of U2 RNAs were determined by reverse transcribing immunoprecipitated snRNAs in the presence of dideoxynucleotides, using a 5′-^{32}P-labeled oligodeoxynucleotide primer complementary to positions 47–62 in *Arabidopsis* U2 RNA. Sequence reactions were performed as described in the Amersham M13 Sequencing Manual, except that the reverse transcriptase was used and additional amounts of nucleotides (50 μM each) were added to all reactions. When the 5′ end of the *Arabidopsis* U2 RNA was to be determined, the dideoxynucleotides were omitted.

Electroporation of plant protoplasts

Suspension cultures of *N.tabacum nia115* (obtained from I.Potrykus, ETH, Zürich) and *O.violaceus* (from C.Matsui, Nagoya University, Japan) were grown in AA (Müller and Grafe, 1978) and MS (Murashige and Skoog, 1962) medium, respectively. *N.tabacum* protoplasts were isolated from 4-day-old cultures by incubation of cells in a solution containing 2.5 mM MES, pH 5.7, 0.6 M sucrose, 2% cellulase R-10, 1% macerozyme R-10 and 0.5% drisellase. After 3 h at 26°C the protoplasts were filtered through a 100 μm sieve. The sieve was washed with 0.2 vol. of a solution containing 1 mM MES, pH 5.7, 10 μM KI, 10 mM $CaCl_2$, 1 mM KNO_3, 1 mM $MgSO_4$ and 16% sucrose. The cell suspension was then transferred to 15 ml Corning tubes (5 ml/tube), and overlaid with 7 ml of EP (10 mM Hepes–KOH, pH 7.2, 150 mM NaCl, 5 mM $CaCl_2$, 0.2 M mannitol). After centrifugation at 1000 r.p.m. for 5 min the banded protoplasts were collected, washed twice with EP, and resuspended in EP (3×10^6 protoplasts/ml).

O.violaceus protoplasts were isolated from 4-day-old cultures by incubation of cells in a solution containing 5 mM MES, pH 5.7, 0.4 M mannitol, 0.1% pectolyase Y-23 and 1% cellulase R-10 (C.Matsui, personal communication). After 1 h at 26°C the protoplasts were filtered through 100 and 50 μm sieves, washed and resuspended as indicated above. Electroporation was carried out essentially as described by Fromm *et al.* (1985). A sample of 0.7 ml of protoplasts was mixed with 100 μg of plasmid DNA. After 10 min on ice the mixture was transferred to the electroporation chamber, made from a 1 ml disposable cuvette. Electroporation was carried out by discharging a 820 μF capacitor charged at 200 V (*O.violaceus*) or 300 V (*N.tabacum*). After 10 min on ice the *N.tabacum* protoplasts were diluted with 10 ml of AA medium of Müller and Grafe (1978), modified by addition of 0.45 M sorbitol, 1 mM xylose, 150 mg/l arabinose, 1 M glucose, 0.4 M inositol, 0.1 M sucrose and *O.violaceus* protoplasts with 10 ml of medium A of Kao and Michayluk (1981), containing 0.4 M glucose, 4.5 μM 2,4-dichlorophenoxyacetic acid, 0.5 μM naphthalenacetic acid and 2.2 μM benzyladenine. Protoplasts were incubated at 26°C for 24–36 h. RNA was prepared from protoplasts according to Chirgwin *et al.* (1979).

Analysis of U2 gene expression by RNase A/T1 mapping

U2 RNA gene expression was analyzed by hybridization of protoplast RNA with ^{32} P-labeled, complementary RNA probes and digestion of hybrids with RNase A and T1 (Melton *et al.*, 1984). Hybridization reactions (10 μl), containing 7.5 μg of RNA and ~1.5 fmol (50 000 c.p.m.) of ^{32}P-labeled probe were carried out in 40 mM Pipes, pH 6.7, 80% formamide, 0.4 M NaCl and 1 mM EDTA. Samples were heated for 5 min at 95°C and then incubated overnight at 50°C. Following the hybridization, 50 μl of a solution containing 10 mM Tris–HCl, pH 7.5, 5 mM EDTA, 0.3 M NaCl, 10 mM $MgCl_2$ and 0.2 U of RNase H (Gibco-BRL) was added. Treatment with RNase H was performed to eliminate potential hybrids with traces of contaminating plasmid DNA. After digestion for 30 min at 37°C, 50 μl of RNase A/T1 mix, containing 80 μg/ml of RNase A and 4 μg/ml of RNase T1 in 10 mM Tris–HCl, pH 7.5, 5 mM EDTA and 0.3 M NaCl were added and incubation continued for 40 min at 26°C. The reaction was stopped by addition of 3 μl of proteinase K (10 mg/ml) and 1 μl of 20% SDS. After 15 min at 37°C the RNA was extracted with phenol/chloroform and analyzed on a 6% polyacrylamide/8 M urea gel. Autoradiography was done at −70°C, with an intensifying screen.

Acnowledgements

We thank K.Wiebauer and G.Goodall for many helpful discussions, B.Hemmings and Y.Nagamine for reading the manuscript, C.Matsui for providing us with *O.violaceus* line, R.Lührmann for antibodies, and D.Edoh for assistance with the deletion analysis.

References

Alonso, A, Beck,E., Jorcano,J.L. and Hovemann,B. (1984) *Nucleic Acids Res.*, **12**, 9543–9550.
Ares,M.,Jr (1986) *Cell*, **47**, 49–59.
Barta,A., Sommergruber,K., Thompson,D., Hartmuth,K., Matzke,M.A. and Matzke,A.J.M. (1986) *Plant Mol. Biol.*, **6**, 347–357.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–185.
Black,D.L., Chabot,B. and Steitz,J.A. (1985) *Cell*, **42**, 737–750.
Branlant,C., Krol,A., Ebel,J., Lazar,E., Haendler,B. and Jacob,M. (1982) *EMBO J.*, **1**, 1259–1263.
Brunel,C., Sri-Widala,J. and Jenateur,P. (1985) In Hahn,F.E. (ed.), *Progress In Molecular and Subcellular Biology*. Springer, Berlin, Vol. 9, pp. 1–52.
Chabot,B., Black,D.L., LeMaster,D.M. and Steitz,J.A. (1985) *Science*, **230**, 1344–1349.
Chirgwin,J.M., Przybyla,E.A., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Functions of Small Major and Minor Nuclear Ribonucleoprotein Particles*. Springer, New York, in press.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Forbes,D.J., Kirschner,M.W., Caput,D., Dahlberg,J.E. and Lund,E. (1984) *Cell*, **38**, 681–689.
Frischauf,A.M., Lehrach,H., Pouska,A., Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.
Fromm,M., Taylor,L.P. and Walbot,V. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5824–5828.
Gram-Jensen,E., Hellung-Larsen,P. and Frederiksen,S. (1979) *Nucleic Acids Res.*, **6**, 321–330.
Henikoff,S. (1984) *Gene*, **28**, 351–359.
Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249–258.
Joshi,C.P. (1987) *Nucleic Acids Res.*, **16**, 6643–6653.
Kao,K.N. and Michayluk,M.R. (1981) *In vitro*, **17**, 645–648.
Kazmaier,M., Tebb,G. and Mattaj,I.W. (1987) *EMBO J.*, **6**, 3071–3078.
Keith,B. and Chua,N.-H. (1986) *EMBO J.*, **5**, 2419–2425.
Keller,E.B. and Noon,W.A. (1985) *Nucleic Acids Res.*, **13**, 4971–4981.
Kiss,T., Toth,M. and Solymosy,F. (1985) *Eur. J. Biochem.*, **152**, 259–266.
Kiss,T., Antal,M. and Solymosy,F. (1987) *Nucleic Acids Res.*, **15**, 1332.
Konarska,M., Filipowicz,W., Domdey,H. and Gross,H.J. (1981) *Nature*, **229**, 112–116.
Krol,A., Ebel,J., Rinke,J. and Luhrmann,R. (1983) *Nucleic Acids Res.*, **11**, 8583–8594.
Leutwiller,L.W., Hough-Evans,B.R. and Meyerowitz,E.M. (1984) *Mol. Gen. Genet.*, **194**, 15–23.
Leutwiller,L.S., Meyerowitz,E.M. and Tobin,E.M. (1986) *Nucleic Acids Res.*, **14**, 4051–4064.
Lührmann,R., Appel,B., Bringmann,P., Rinke,J., Reuter,R., Rothe,S. and Bald,R. (1982) *Nucleic Acids Res.*, **10**, 7103–7113.
Lund,E., Kahn,B. and Dahlberg,J.E. (1985) *Science*, **229**, 1271–1274.
Lund,E. and Dahlberg,E.F. (1987) *Genes Devel.*, **1**, 39–46.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maniatis,T. and Reed,R. (1987) *Nature*, **325**, 673–678.
Mattaj,I.W. (1986) *Cell*, **46**, 905–911.
Mattaj,I.W. and Zeller,R. (1983) *EMBO J.*, **2**, 1883–1891.
Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) *Protein Engineering*, **1**, 67–74.
Melton,D.A., Krieg,P.A., Rebagliatti,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Mount,S.M., Petterson,I., Hinterberger,M., Karmas,A. and Steitz,J.A. (1983) *Cell*, **33**, 509–518.
Müller,A.J. and Grafe,R. (1978) *Mol. Gen. Genet.*, **161**, 67–76.
Murashige,T. and Skoog,F. (1962) *Physiol. Plantarum*, **15**, 473–497.
Murphy,J.T., Burgess,R.R., Dahlberg,J.E. and Lund,E. (1982) *Cell*, **29**, 265–274.
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.
Murray,M.G. and Thompson,W.F. (1980) *Nucleic Acids Res.*, **8**, 4321–4325.

Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
Reddy,R. (1986) *Nucleic Acids Res.*, **14**, r61–r72.
Short,K.C. and Torrey,J.G. (1972) *Plant Physiol.*, **49**, 155–160.
Skuzeski,J.M. and Jendrisak,J.J. (1985) *Plant Mol. Biol.*, **4**, 181–193.
Strub,K., Galli,G., Busslinger,M. and Birnstiel,M.L. (1984) *EMBO J.*, **3**, 2801–2807.
Tani,T., Watanabe-Nagasu,N., Okada,N. and Oshima,Y. (1983) *J. Mol. Biol.*, **168**, 579–594.

Received on December 21, 1987

Note added in proof

The manuscript referred to as 'K.Wiebauer, J.Herrero and W.Filipowicz, submitted' will appear in *Molecular and Cellular Biology*. The sequences of U2 RNA genes are being deposited at the EMBL Data Library with accession numbers X06473 to X06478.