U6 snRNA genes of *Arabidopsis* are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase II-transcribed U-snRNA genes

Franz Waibel and Witold Filipowicz*
Friedrich Miescher-Institut, PO Box 2543, 4002 Basel, Switzerland

Received April 20, 1990; Accepted May 23, 1990

EMBL accession nos X52527, X52528, X52529

ABSTRACT

Previously we have demonstrated that the U2 snRNA genes from the higher plant *Arabidopsis thaliana* contain two upstream elements, the USE with sequence RTCCCACATCG and a −30 'TATA' box, which are essential for transcription by RNA polymerase II, and that the conserved spacing of about four helical DNA turns between these elements is important for optimal promoter function. We have now isolated three genes encoding U6 RNA in *Arabidopsis*. Transcription of these genes in transfected protoplasts of *Nicotiana plumbaginifolia* is resistant to α-amanitin indicating that they are transcribed by RNA polymerase III. The upstream regions of three *Arabidopsis* U6 genes contain USE and −30 TATA-like elements similar to those found to be important for transcription of U2 RNA genes but the spacing between the two elements is about 10 bp closer than in the U2 genes. Using synthetic U6 genes we demonstrate that the USE and TATA elements are indispensable for their transcription, the TATA boxes of U2 and U6 genes are interchangeable, and that the intragenic A box-like sequence of U6 gene is not essential. Increasing the distance between the USE and TATA by 10 bp inactivates U6 gene transcription, demonstrating that proper positioning of the elements is also important for transcription by RNA polymerase III. The data indicate that the structure of U-snRNA gene promoters and the determinants of polymerase specificity are completely different between vertebrates and plants.

INTRODUCTION

In the nuclei of eukaryotic cells RNA polymerase II (pol II) transcribes all protein-coding genes and most of the genes encoding U-snRNAs, the small nuclear RNAs involved in RNA processing. RNA polymerase III (pol III) synthesizes tRNAs, 5S rRNA and other types of cytoplasmic and nuclear small RNAs, including U6 snRNA (reviewed in refs 1−4). Accurate initiation by either of these enzymes requires several protein factors, most of which interact with specific DNA sequences within promoters (2,3). For a long time it has been recognized that class II and class III genes differ not only in the structure of their promoter elements, but also, and more importantly, in the location of these elements within a gene. The promoter elements required for initiation by pol II are generally situated upstream of the coding region. On the other hand, the promoters of pol III genes are usually located internally, within the coding region; tRNA and 5S rRNA genes containing internal A and B (or C) boxes are examples of such genes (2,5,6). In recent years, however, this distinction between pol II and pol III promoters has become much less rigid. (i) It has been found that efficient transcription of some tRNA and 5S rRNA genes (reviewed in ref. 2) and a 7SL gene (7), depends upon upstream sequences, although these sequences and the factors interacting with them are not well characterized. (ii) The Epstein-Barr virus-encoded small RNA (EBER) genes (8) are transcribed by pol III but contain functional promoter elements typical of both pol III genes (intragenic A and B boxes) and pol II genes (ATF- and Sp1-like, and possibly TATA-like, upstream elements). (iii) Transcription from the human c-*myc* gene promoter, in addition to being initiated by pol II, may also be initiated accurately, although very inefficently, by pol III (9,10). The physiological significance of this finding is not clear but transcription by pol III appears to require the TATA box and no intragenic sequences (10). (iv) Genes encoding U6 (11−16) and 7SK (17,18) RNAs in vertebrates are transcribed by pol III but do not require any intragenic elements. Instead, their activity depends upon three upstream signals: the Distal and Proximal Sequence Elements (DSE and PSE), and an AT-rich box which resembles the TATA box found in mRNA genes. The first two elements are structurally and functionally identical to the elements essential for transcription of the pol II-specific snRNA-genes (reviewed in ref. 4). The AT-rich box is unique to the pol III genes and is the dominant signal conferring pol III specificity to the promoter (12,16).

Previously we have demonstrated that the pol II-transcribed U2 genes of *Arabidopsis* contain two essential promoter elements, a −30 TATA element which is structurally and functionally indistinguishable from the TATA boxes of mRNA coding genes,

* To whom correspondence should be addressed

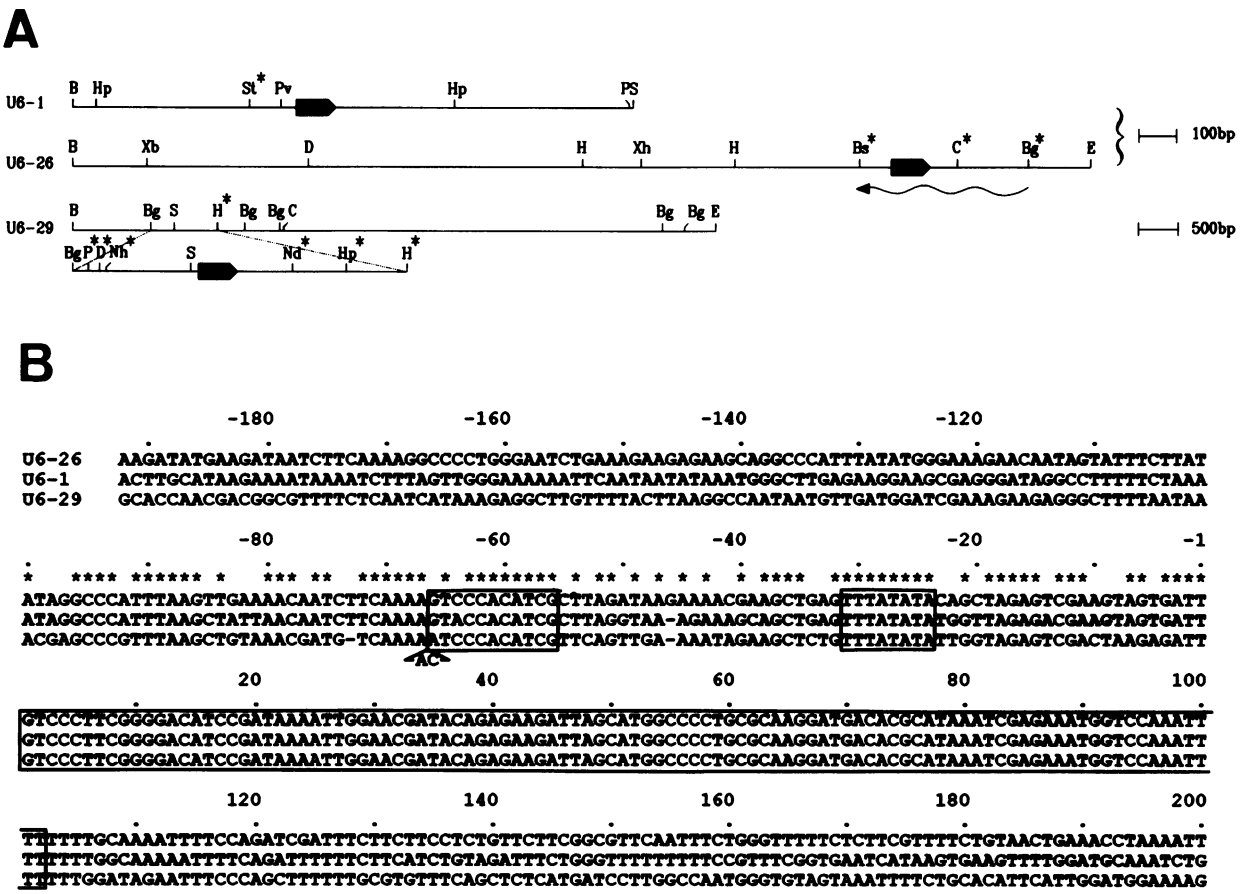


Figure 1. Structure of three *Arabidopsis* U6 RNA genes. (A) Restriction maps of the 1.4−8.2 kb DNA fragments containing three different genes. Arrow shaped boxes indicate coding regions. B, Bg, Bs, C, D, E, H, Hp, Nd, Nh, P, Pv, S, St, Xb and Xh are *Bam*HI, *Bgl*II, *Bst*XI, *Cla*I, *Dra*I, *Eco*RI, *Hin*dIII, *Hpa*I, *Nde*I, *Nhe*I, *Pst*I, *Pvu*II, *Sal*I, *Stu*I, *Xba*I and *Xho*I sites, respectively. Restriction sites that are not unique are identified by asterisks. The wavy line below the U6-26 gene indicates the DNA fragment that was transcribed to yield antisense RNA probe. (B) Sequences of the U6 genes. Numbering of the upstream non-coding sequence corresponds to the gene U6-26. Other sequences are aligned with U6-26 to accentuate similarities in the −100/−1 region; further upstream no significant similarity is found. Asterisks indicate nucleotide identity in the three promoter regions. Two sequence motifs, RT(C/A)CCACATCG (pos. −66/−56) and TTTATATA (−31/−24) which are similar or identical to the promoter elements of the *Arabidopsis* U2 and U5 genes (20,21) are boxed, as are the coding regions.

and an Upstream Sequence Element (USE), RTCCCACATCG, which is found in U-snRNA, but not in mRNA, genes of plants (19). In plant genes encoding the 2,2,7-trimethylguanosine (m_3G)-capped U-snRNAs (U1, U2, U4 and U5; 20−25 and our unpublished results), these two elements are centered approximately four DNA helical turns apart, a property required for optimal promoter function (19). In this report we show that the pol III-transcribed U6 RNA genes of *Arabidopsis* also contain both the USE and the TATA-like elements, but that these elements are spaced one helical turn closer than in pol II-transcribed U-snRNA genes. The USE and TATA-like upstream motifs of the U6 genes appear to be the only essential elements for their transcription.

MATERIALS AND METHODS

Isolation of the U6 RNA genes

Unless indicated otherwise, all techniques used for manipulating DNA were as described by Maniatis *et al.* (26). A genomic library, in the λEMBL3 vector, of *Arabidopsis thaliana* strain Landsberg was screened as described (20), using as a probe a 5′-^{32}P-labeled 29-mer oligonucleotide complementary to pos. 39−67 of *Arabidopsis* U6 RNA. Filters were washed with 1×SSC, 0.1% SDS for 30 min at 65°C before autoradiography. Fragments of purified phage DNA which hybridized to the probe were subcloned to pBluescript(+) (Stratagene): the 1.4 kb *Bam*HI-*Sal*I fragment of λU6-1 to yield pU6-1; the 2.6 kb *Bam*HI-*Eco*RI fragment of λU6-26 to yield pU6-26large; and the 8.2 kb *Bam*HI-*Eco*RI fragment of λU6-29 to yield pU6-29. The 0.85 kb *Hin*dIII-*Eco*RI subfragment from pU6-26large was further subcloned to pGEM2 (Promega) to yield pU6-26. Approximately 0.9 kb of each insert was sequenced on both strands by the dideoxynucleotide method using double-stranded DNA templates and universal or internal primers. Complete sequences are deposited in the EMBL, Genbank and DDBJ Nucleic Sequence Databases (accession numbers X52527,X52528,X52529). Southern blots with genomic DNA were hybridized and washed as described (20) using the *Sal*I-*Sph*I fragment of pU6syn as a probe.

5′ end mapping of U6 RNA by reverse transcription

This was done by primer extension using as a template low molecular weight RNA isolated from *Arabidopsis* plants (20). A 29 nt-long oligodeoxynucleotide, complementary to positions 39−67 of *Arabidopsis* U6 RNA, served as a primer for reverse transcription (20).

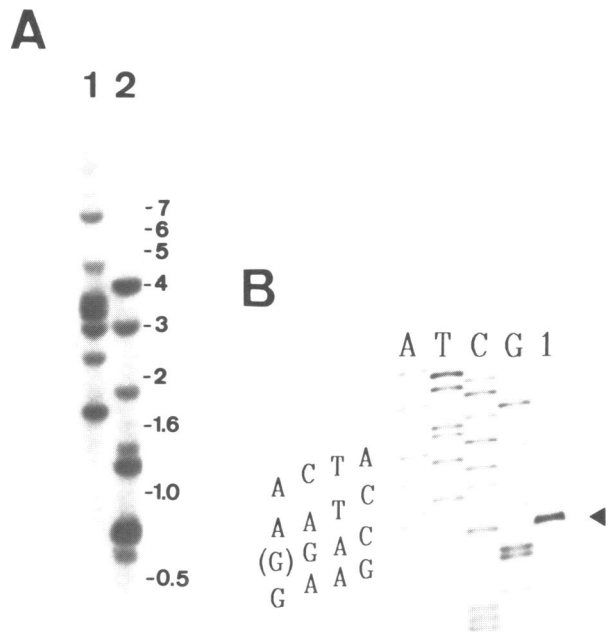


Figure 2. Southern blot analysis (A) and determination of the 5′ end of *Arabidopsis* U6 RNA by primer extension (B). (A) 5 μg of *Arabidopsis* genomic DNA was restricted with *Hin*dIII (lane 1) and *Dra*I (lane 2). DNA was probed with the *Sal*I-*Sph*I fragment of pU6syn labeled by the random priming method. Size markers (in kb) are indicated. (B) The product of reverse transcription (lane 1) is indicated by an arrow. Lanes A, T, C and G show a sequence ladder of the gene U6-26 (plasmid pU6-26, coding strand). Due to band compression, one G-specific band is not evident and is shown in brackets.

Transient expression in transfected protoplasts and RNase A/T_1 mapping

Transient expression was tested in mesophyll protoplasts of *Nicotiana plumbaginifolia*, which were transfected using the polyethylene glycol method (27). Unless indicated otherwise, 10 μg of each plasmid, corresponding to a subsaturating amount (see Fig. 5D), was used per transfection of 6×10^5 protoplasts. α-Amanitin was added at the time of transfection and subsequently diluted with the culture medium to yield the concentration as indicated. RNA was prepared 24 h after transfection and used for RNase A/T_1 mapping (28). The U6 and U2 gene-specific antisense RNA probes (440 and 510 nt-long, respectively) were synthesized *in vitro* by SP6 polymerase using [α-^{32}P]GTP (sp. act. 160 Ci/mmol) as the label and pU6-26 linearized with *Bst*XI (see Fig. 1A) or pGEM2.U2.2 linearized with *Hin*dIII (20) as templates. Probes were purified on polyacrylamide gels. The U6-26 probe was used in all experiments, except those with the U6synAbox mutants (Fig. 6). For those the cognate complementary probes were transcribed from the respective plasmids (pU6synAbox1 through Abox3) linearized with *Bam*HI. 3 μg of RNA and 6 fmoles (100−150,000 Cerenkov cpm) of each probe was used per mapping reaction. Protected fragments specific for expressed *Arabidopsis* U6 and U2 RNAs, and respective endogenous fragments (used for normalization of transfection efficiency and RNA recovery) were quantitated by scintillation counting (27). Similar results were obtained irrespective of whether only the 105 nt bands or the 105 nt plus 93−97 nt bands were used for quantitation.

Construction of pU6syn and its derivatives

The 188 bp-long U6syn sequence extending from the *Bam*HI to the *Sph*I site was assembled from overlapping oligonucleotides (the shortest overlap being 13 bp) ranging in size between 21 nt and 44 nt, kindly supplied by J.Jiricny and W.Zürcher of our Institute. Eight internal 5′-phosphorylated and two terminal non-phosphorylated oligonucleotides were annealed (28), ligated by T_4 DNA ligase and inserted into pTZ19R precut with *Bam*HI and *Sph*I, yielding pU6syn. pU6synA and pU6synB were constructed by cloning the *Sal*I-*Ava*II or *Sal*I-*Hin*dIII fragments of pU6-29 into the corresponding sites of pU6syn (Fig. 5B).

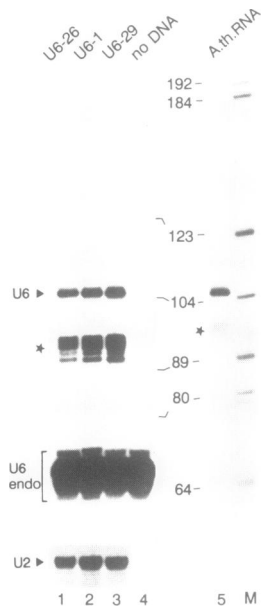
Construction of U6syn mutants

The construct pU6synΔUSE was prepared by cleavage of pU6syn with *Bam*HI and *Xho*I, followed by blunting of ends by mung bean nuclease and ligation. Constructs pU6syn-TATA1 and -TATA2 were generated by inserting appropriate linkers, containing mutations in the TATA-like regions, into the *Nco*I/*Sal*I-restricted pU6syn. Mutant pU6syn+10 was obtained by replacement of the *Xho*I-*Nco*I fragment in pU6syn with a longer linker containing the sequence CTCGAGTAAAAATA-AGAAAGATTCCATGG. Mutants U6synAbox1, Abox2 and Abox3 were obtained by assembly of appropriate overlapping oligonucleotides as described for the gene U6syn. The identity of all constructs was verified by sequence analysis.

RESULTS

Structure of Arabidopsis U6 genes

Three different genes encoding U6 RNA in *Arabidopsis* were isolated and sequenced (Fig. 1). Southern blot analysis (Fig. 2A) indicated the presence of 3 or 4 additional sequences hybridizing to the U6 gene-specific probe. The 5′ border of the coding regions was assigned to G at pos. +1 by primer extension analysis of *Arabidopsis* U6 RNA (Fig. 2B). Assignment of the 3′ end,



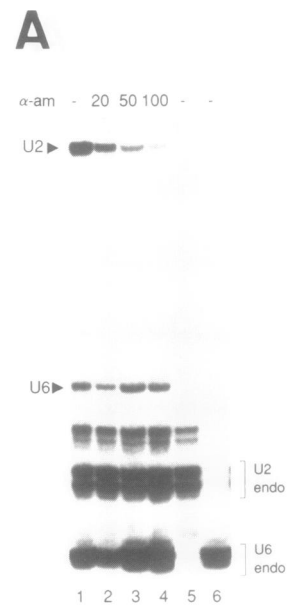


Figure 3. Transient expression of *Arabidopsis* U6 genes in protoplasts of *Nicotiana plumbaginifolia*. Equimolar amounts of plasmids bearing different U6 genes were used; pUC18 DNA was added to maintain a total of 20μg of DNA in each transfection. In addition, 5 μg of plasmid pU2.2 (20) containing the *Arabidopsis* U2 gene was cotransfected to provide an internal standard for normalization of gene expression. RNA from protoplasts (lanes 1–4) or from *Arabidopsis* plants (lane 5) was analyzed by RNase A/T_1 mapping using the U6 (upper panel) or U2 (lower panel) RNA probes. Lane 1, 8 μg of pU6-26; lane 2, 9 μg of pU6-1; lane 3, 20 μg of pU6-29; lane 4, no DNA. Lane M, size markers (end-labeled *Hae*III digest of pBR322). Lane 5, mapping of *Arabidopsis* RNA (5 μg). The approximately 105 nt-long protected fragment, marked by an arrow, is longer than U6 RNA (102 nt) because the stretch of A residues in the probe adjacent to the coding region is resistant to RNases. The 93–97 nt-long bands marked with asterisks and present at variable levels in different mapping experiments result, most likely, from the partial digestion of the RNA hybrid by RNase A at the AU-rich 3′ end. Fragments protected by endogenous *N. plumbaginifolia* U6 RNA are bracketed.

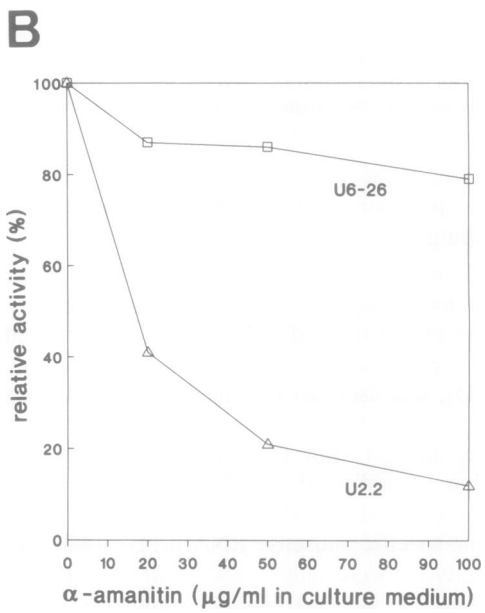


Figure 4. Effect of α-amanitin on transcription of the *Arabidopsis* U6 and U2 genes. (A) Protoplasts were cotransfected with pU6-26 and pU2.2 and incubated in the presence of indicated concentrations of α-amanitin; RNA was mapped with a mixture of U6 and U2 gene-specific probes (lanes 1–4). Lanes 5 and 6, RNA from non-transfected protoplasts, mapped with the U2 and U6 probes, respectively. The endogenous fragments used in quantitation are marked. Because the U-snRNAs are metabolically stable (4), no effect of α-amanitin on endogenous RNA is seen. (B) Quantitation of the inhibition data. Each value is an average from two independent transfections.

located within a T-rich stretch, which usually functions as a pol III termination signal (29), is based on comparison with the broad bean U6 RNA (30) which has 24 nucleotides identical to *Arabidopsis* DNA at its 3′ end. The genes U6-1, U6-26 and U6-29 all encode 102 nt-long RNAs of identical sequence. *Arabidopsis* U6 RNA differs from broad bean U6 RNA (30) by two nucleotide substitutions (C to U in pos. 35 and 78) and the presence of 4 additional nucleotides, GUCC, at the 5′ end. Since the 5′ end of the *Arabidopsis* U6 RNA sequence, but not that of bean, can be folded into the conserved stem/loop structure (31), it is likely that the bean sequence is incomplete at its 5′ end (see also re. 32). Alignment of the *Arabidopsis* U6 RNA sequence with U6 RNAs from other organisms (data not shown), reveals a high degree of sequence conservation, a feature noted previously (31).

Transcription of U6 genes in transfected protoplasts

Activity of the cloned U6 genes was tested by measuring their expression in transfected protoplasts of *Nicotiana plumbaginifolia*. Transcription of a cotransfected *Arabidopsis* U2 gene was also measured to provide an internal standard. RNA was isolated 24 h after transfection and analyzed by RNase A/T1 protection, using ^{32}P-labeled antisense RNA probes specific for the *Arabidopsis* U6 or U2 genes. Transfection with each U6 gene resulted in accumulation of comparable amounts of protected RNA fragments of expected length, indicating that all three genes are active (Fig. 3, lanes 1–3); mapping of U6 RNA isolated from *Arabidopsis* yielded similar protected fragments (lane 5).

U6 genes of *Arabidopsis* are transcribed by RNA polymerase III

To determine whether the *Arabidopsis* U6 RNA is synthesized by pol III, we studied the effect of α-amanitin on transcription.

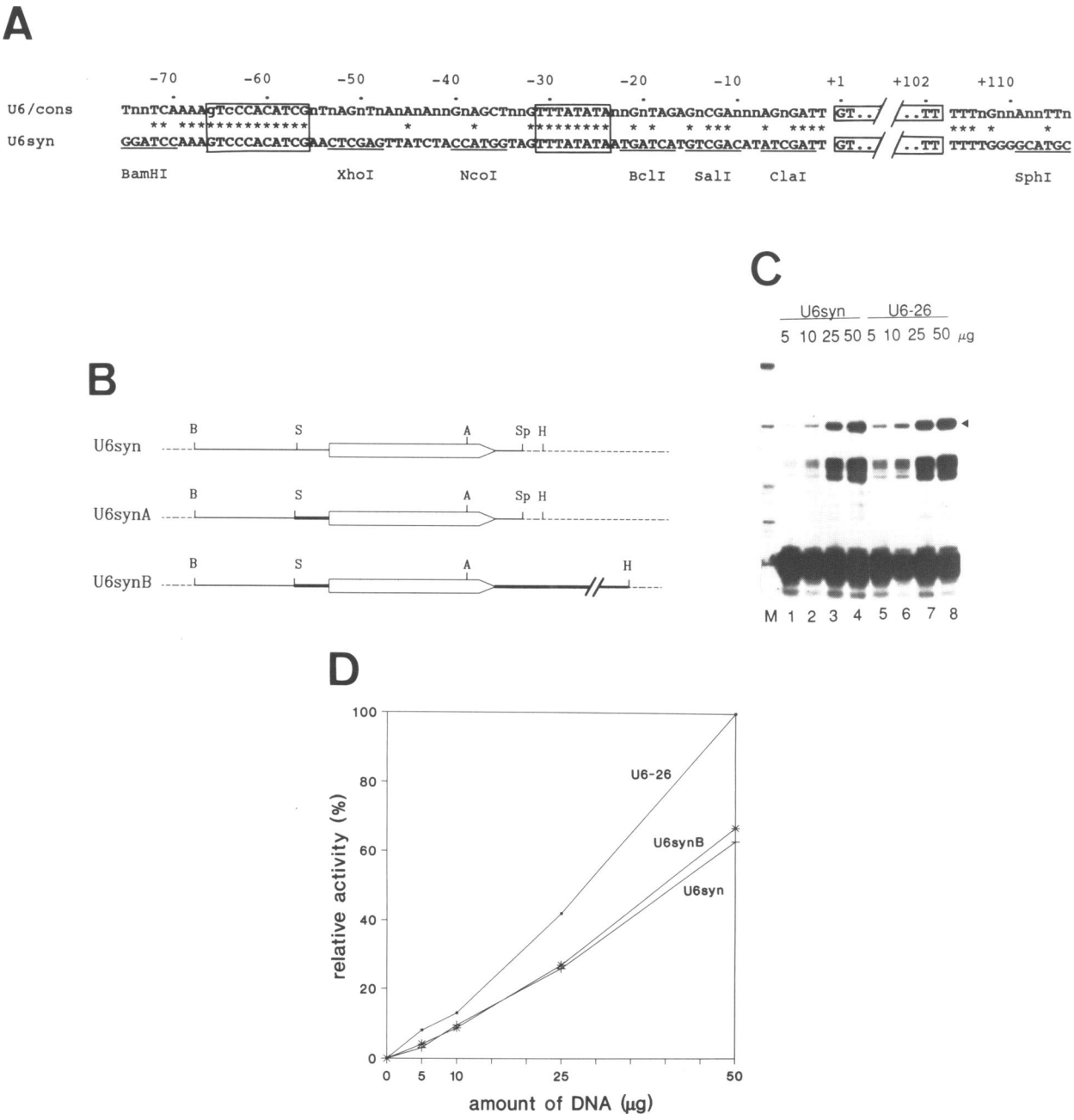


Figure 5. Design and activity of the synthetic U6 gene, U6syn and its variants. (A) Design of the gene U6syn. A U6 gene consensus sequence (U6/cons) is shown in the upper line. Nucleotides conserved between three *Arabidopsis* genes are in upper case. Non-conserved positions are marked with 'n' except for pos. −66 and −64 of the USE (indicated by g or c) in which G or C are present in two out of three genes. Positions in the upstream and downstream regions of U6syn which are identical to the U6 consensus are indicated with asterisks. The coding part (boxed) of U6syn is the same as in the authentic genes. Restriction sites incorporated into U6syn are indicated. (B) Schematic drawing of U6syn and its derivatives. Vector sequence is indicated by a broken line. Sequences derived from gene U6-29 are shown as thick lines. Relevant restriction sites are shown: B, *Bam*HI; S, *Sal*I; A, *Ava*I; Sp, *Sph*I; H, *Hin*dIII. (C) Comparison of activities of the genes U6syn (lanes 1−4) and U6-26 (lanes 5−8) in protoplasts. The amount of plasmid used for transfection is indicated at the top. (D) Quantitation of the results shown in panel C. The data for expression of the gene U6synB, not shown in panel C, are also included.

A concentration of 50−100 µg of the toxin per ml of protoplast culture medium is known to inhibit transcription by pol II but not by pol III (33). Protoplasts were cotransfected with plasmids containing U6 and U2 genes of *Arabidopsis* and incubated in the presence of various concentrations of α-amanitin. The toxin strongly inhibited expression of the pol II-specific U2 gene (and had a similar inhibitory effect on expression of an mRNA-coding gene in protoplasts; data not shown) but had little effect on the synthesis of U6 RNA (Fig. 4). These results, together with the presence of the T-rich terminator-like signal in the U6 genes, and the absence of the m_3G cap in plant U6 RNAs (30; and our unpublished results) establish that in plants, as in vertebrates (34−36) and yeast (37), U6 RNA genes are transcribed by pol III. On the other hand, sensitivity of the synthesis of the m_3G-capped U2 snRNA supports the conclusion that in plants, as in other organisms (4), this gene is transcribed by pol II.

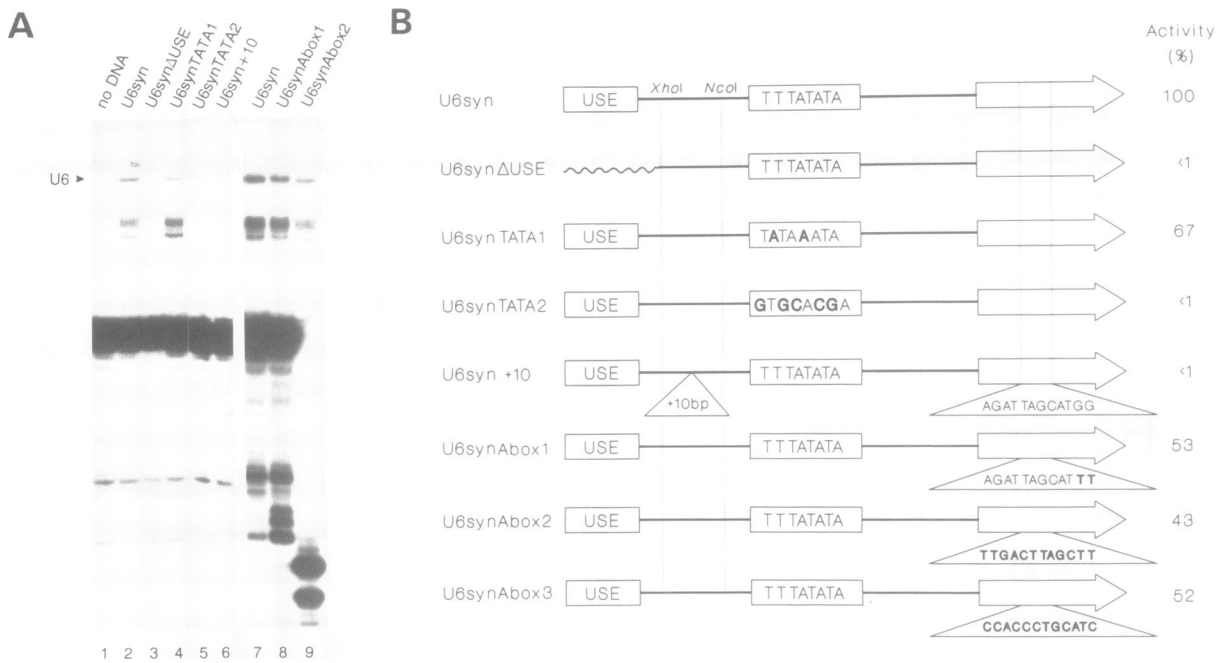


Figure 6. Activity of the U6syn promoter mutants. The structure of the different mutants and their activities, relative to the activity of U6syn, are shown in the diagram (A). The USE and TTTATATA elements are boxed; mutations in the latter are in bold letters. The coding region is represented by an arrow and mutations in the A box-like sequence AGATTAGCATG (pos. 43−54) are indicated. In the Abox3 mutant the sequence AGA, located immediately upstream, is additionally changed to TAA. The relevant restriction sites are marked. The wavy line in U6synΔUSE corresponds to pTZ19R vector sequence. Plasmids used for cotransfection together with pU2.2 are indicated at the top of the autoradiogram (B). RNAs analyzed in lanes 7−9 were isolated 8 hr after transfection. Mapping was performed with the U6-26 probe with the exception of RNAs transcribed from the U6synAbox mutants which were mapped with cognate complementary probes. Lane 1, control without cotransfected U6 gene, probed with the U6-26 gene-specific probe. Quantitative data are averages from three independent transfections. Mapping of RNA expressed from U6synAbox3 is not shown. Each value was normalized for variation in transfection efficiency using U2 gene expression as an internal standard (not shown).

Assembly and activity of synthetic U6 genes

The three *Arabidopsis* U6 genes described in this work contain several stretches of conserved sequences in the non-coding upstream regions (Fig. 1B and 5A). Most notably, these regions contain two conserved elements similar or identical to the −30 TATA-like box and the USE which are sufficient for transcription of *Arabidopsis* U2 RNA genes (19). The U6 gene counterparts have the sequences TTTATATA (pos. −30/−23) and RTCCCACATCG (pos. −66/-56). In promoters of U2 genes and other genes encoding m_3G-capped U-snRNAs (U1, U4, U5), the USE and the TATA-like element are centered approximately four helical turns apart (21−25); this property is important for optimal promoter function (19). In the cloned pol III-transcribed U6 genes the USE is positioned approximately 10 bp closer to the TATA-like sequence.

To define the promoter signals required for U6 gene transcription we have constructed a synthetic gene, called U6syn (Fig. 5A). The 70 bp-long upstream region of U6syn contains the USE and TTTATATA elements at the conserved positions. The remaining upstream sequence is mostly unrelated to that of the three *Arabidopsis* U6 genes (see U6 consensus in Fig. 5A) and contains several restriction sites in order to facilitate manipulation of DNA. The coding region of U6syn is identical to the wild-type sequence and a stretch of T residues is included downstream to provide a termination signal. Two U6syn variants were also constructed (Fig. 5B). The variant U6synA contains 15 bp of the authentic U6-29 upstream sequence replacing U6syn region −15/−1. The U6synB contains 430 bp of the U6-29 downstream sequence (pos. 103/533) fused to the coding region of U6synA. The latter variant was made in order to test whether sequences positioned downstream of the T stretch are important for gene expression. All three U6syn genes were transcribed with similar efficiency, their activity corresponding to 60−65% of that of the authentic U6 RNA gene (Fig. 5C and D, and data not shown). Sequences included in the gene U6syn, containing the USE and TATA upstream boxes and the 3′ end T-rich terminator were therefore sufficient to drive efficient synthesis of U6 RNA.

Activity of the U6syn mutants

Several mutants of U6syn were tested (Fig. 6). Deletion of the USE (mutant U6synΔUSE) or replacement of the TATA-like box (TTTATATA) by an unrelated sequence (mutant U6synTATA2) completely abolished transcription, indicating that these elements are absolutely essential as was previously demonstrated for the pol II-specific U2 snRNA gene (19). Replacement of the TTTATATA by the TATA-like motif TATAAATA, conserved in the *Arabidopsis* U2 genes (20), resulted in only a small decrease in activity (mutant U6synTATA1). The TATA sequences of the U6 and U2 genes, TTTATATA and TATAAATA respectively, are therefore interchangeable. In the mutant U6syn+10, the distance between the USE and TTTATATA was increased by 10 bp, resulting in the element spacing characteristic of the pol II promoters. This modification completely abolished RNA synthesis indicating that the conserved spacing between the USE and TATA elements, corresponding to approximately three helical DNA turns, is essential for transcription by pol III.

The U6 genes in *Arabidopsis* and other organisms contain the sequence AAGATTAGCATGG (pos. 42−54; Fig. 1B) which differs in only two positions (A_{42} and A_{45}) from the intragenic A box consensus of tRNA genes (2). This sequence is part of the highly conserved region of U6 RNA which base-pairs with U4 RNA in the U4/U6 snRNP complex (31). Three mutants were constructed to test the importance of the A box-like sequence for transcription. U6synAbox1 has two additional positions of the consensus, G_{53} and G_{54}, replaced by T residues, and in U6synAbox2 and U6synAbox3 the sequence AGATTAGCATGG is replaced by different unrelated sequences. All mutations resulted in only about a 2-fold decrease in the amount of accumulated U6 RNA (Fig. 6 and data not shown). We have not investigated whether this is due to decreased synthesis or stability of modified U6 RNAs.

DISCUSSION

We have isolated three active genes encoding U6 snRNA in *Arabidopsis thaliana*. Sensitivity of their transcription to α-amanitin, the presence of T-rich terminator-like sequences at their 3′ ends, and lack of the m_3G cap in plant U6 RNA (30; and our unpublished results) indicate that U6 genes in higher plants are, as in vertebrates (34−36) and yeast (37), transcribed by pol III. Despite being transcribed by a different RNA polymerase, the U6 genes of *Arabidopsis* contain two upstream elements, the USE and TATA, which are similar to the upstream *cis* elements sufficient for transcription of the pol II-specific U snRNA genes of plants (21−25). Experiments with synthetic U6 genes indicated that the USE and TATA-like elements are also indispensable for transcription by pol III.

Our findings that TATA-like boxes are present in both pol II- and pol III-specific snRNA genes of *Arabidopsis*, and that these elements are exchangeable between the two classes of promoters (19; and this work), offer a unique opportunity to address the question whether the same or different TATA factors are involved in transcription by pol II and pol III. Upstream TATA-like elements have been found in a number of pol III genes (2 [and references herein],8,11,14,15,17,38) although they do not always closely resemble the TATA consensus of mRNA genes. The TATA box is also essential for initiation of c-*myc* gene transcription by pol III but the nature of the protein factor involved has not been established (10). We are presently testing whether transcription of the U2 and U6 genes that contain identical TATA sequences (genes U2syn and U6synTATA1) is similarly affected by different mutations in this region.

Another essential promoter element, found in the pol II and pol III-specific U-snRNA genes of plants is the USE. This element is snRNA gene-specific; it is not found in mRNA or any other genes of plants (21). In all class II U-snRNA genes characterized to date the USE is centered approximately four helical DNA turns upstream of the TATA box (21−25). This location is essential for the promoter function since increase of the USE-TATA distance by 50 bp or its shortening by 10 or 20 bp decreases transcription of the U2 gene by 10−20 fold (19). In the three isolated U6 genes of *Arabidopsis* the spacing between the USE and TATA is also highly conserved but it corresponds to about three helical DNA turns in contrast to the four turns of the class II promoters. Two additional U6 gene sequences of *Arabidopsis* characterized by PCR amplification (39) and the U6 gene isolated recently from tomato (32) also contain the USE and TATA elements centered approximately three helical turns apart. This spacing is clearly essential for transcription by pol III since the insertion of 10 bp of DNA into the USE-TATA spacer of the U6syn inactivates the gene. Recently, using several different suitably manipulated pol II/pol III gene hybrids, we have demonstrated that a single helical turn difference in the spacing of USE and TATA elements is indeed the major determinant of RNA polymerase specificity during transcription of U-snRNA genes of *Arabidopsis*. It was possible to achieve conversion of the pol III promoter to a pol II-specific promoter, and vice versa, by changing the distance between the USE and TATA elements (F.W. and W.F., manuscript submitted).

Although the *Arabidopsis* U6 genes contain a sequence similar to an A box, which is the only U6 intragenic region resembling a known pol III-specific transcription element, this sequence can be modified without strongly affecting transcription. In vertebrates too, synthesis of U6 RNA does not depend upon the A box or other intragenic sequences (11,14,16,40). However, our experiments have not addressed the question whether sequences at the transcription start site are important for the promoter activity. Transcription of all plant U-snRNA genes (21−25) starts at a purine residue preceded by a pyrimidine. Transcription of the U6 gene of *Xenopus* shows similar requirements (12).

The structure of U-snRNA gene promoters and the determinants of polymerase specificity are different in plant and vertebrate cells. Promoters of the pol II- and pol III-specific U-RNA genes in vertebrates contain two common upstream elements, DSE and PSE, positioned in the −250 and −50 regions, respectively (reviewed in refs 4,41). The DSE is an enhancer-like element. It is usually composed of several sequence motifs, such as octamer- or SP1-binding sites, similar to the ones functioning in vertebrate mRNA genes. The PSE element in pol II U-snRNA genes determines selection of the correct start site; its function is therefore equivalent to that of the TATA box in mRNA genes. The DSE and PSE are sufficient for transcription of the U2 gene by pol II. The U6 genes contain an additional element, an AT-rich box located in the −30 region, which confers pol III specificity (12,16). The similarity of class II and III promoters in vertebrates is supported by the demonstration that although U6 RNA is a product of pol III, a low level of transcription initiation by pol II also occurs (12). Furthermore, it has been possible to convert a pol II-specific U2 promoter into a pol III promoter or a U6 promoter to a pol II promoter by either inserting or mutating the AT-rich element (16).

Both classes of plant U-snRNA gene promoters resemble eubacterial promoters, which also contain two conserved upstream sequences, the −35 TTGACA and −10 TATAAT regions (42), the precise spacing of which is important for optimal promoter function (43,44; for a discussion of the similarity of vertebrate 7SK and U6 promoters with eubacterial promoters, see also refs 17,45). Furthermore, −30 TATA-like elements, similar to those found in plant U-snRNA genes and also in pol II mRNA genes, are present in archaebacterial genes (46,47). It is likely that all these promoters have evolved from a common ancestor and that one mechanism of promoter diversification in eukaryotes involved changes in the spacing between different promoter elements recognized by the transcriptional apparatus.

The structural relationships between different promoters discussed above and the sequence similarities between the large subunits of three eukaryotic polymerases, and between eukaryotic and prokaryotic polymerase subunits (48,49), all argue for a common ancestry of the eukaryotic and prokaryotic transcriptional apparatus. Protein sequence analysis has revealed

(48) that pol II and pol III enzymes are more closely related to each other than to pol I and may have evolved from a common progenitor more recently than pol I. Our results and the similarity of some other pol II and pol III promoters strongly support this notion.

ACKNOWLEDGEMENTS

We thank the members of our laboratory and Karin Wiebauer for many helpful discussions, Barbara Hohn, Greg Goodall and Jean-Pierre Jost for critical comments on the manuscript, and Jirina Petruska for excellent assistance in protoplast transfections.

REFERENCES

1. Sentenac, A. (1985) CRC Crit. Rev. Biochem.,**18**, 31–90.
2. Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) Ann. Rev. Biochem., **57**, 873–914.
3. Wasylyk, B. (1988) CRC Crit. Rev. Biochem.,**23**, 77–120.
4. Dahlberg, J.E. and Lund, E. (1988) In Birnstiel, M.L.(ed.) *Structure and Function of Small Nuclear Ribonucleoprotein Particles*. Springer Verlag, Heidelberg, pp. 38–70.
5. Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980) Cell, **19**, 13–25.
6. Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Nature, **294**, 626–631.
7. Ullu, E. and Weiner, A.M. (1985) Nature, **318**, 371–374.
8. Howe, J.G. and Shu, M.-D. (1989) Cell, **57**, 825–834.
9. Chung, J., Sussman, D.J., Zeller, R. and Leder, P. (1987) Cell, **51**, 1001–1008.
10. Bentley, D.L., Brown, W.L. and Groudine, M. (1989) Genes Dev., **3**, 1179–1189.
11. Carbon, P., Murgo, S., Ebel, J.-P., Krol, A., Tebb, G. and Mattaj, I.W. (1987) Cell, **51**, 71–79.
12. Mattaj, I.W., Dathan, N.A., Parry, H.D., Carbon, P. and Krol, A. (1988) Cell, **55**, 435–442.
13. Bark, C., Weller, P., Zabielski, J., Janson, L. and Pettersson, U. (1987) Nature, **328**, 356–359.
14. Das, G., Henning, D., Wright, D. and Reddy, R. (1988) EMBO J.,**7**, 503–512.
15. Kunkel, G.R. and Pederson, T. (1988) Genes Dev. , **2**, 196–204.
16. Lobo, S.M. and Hernandez, N. (1989) Cell, **58**, 55–67.
17. Murphy, S., Di Liegro, C. and Melli, M. (1987) Cell, **51**, 81–87.
18. Kleinert, H. and Benecke, B.-J. (1988) Nucleic Acids Res., **16**, 1319–1331.
19. Vankan, P. and Filipowicz, W. (1989) EMBO J., **8**, 3875–3882.
20. Vankan, P. and Filipowicz, W. (1988) EMBO J., **7**, 791–799.
21. Vankan, P., Edoh, D. and Filipowicz, W. (1988) Nucleic Acids Res., **16**, 10425–10439.
22. van Santen, V.L., Swain, W. and Spritz, R.A. (1988) Nucleic Acids Res., **16**, 4176.
23. Abel, S., Kiss, T. and Solymosy, F. (1989) Nucleic Acids Res., **17**, 6319–6337.
24. Brown, J.W.S. and Waugh, R. (1989) Nucleic Acids Res., **17**, 8991–9001
25. Hanley, B.A. and Schuler, M.A. (1989) Nucleic Acids Res., **17**, 10106.
26. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Goodall, G., Wiebauer, K. and Filipowicz, W. (1990) Meth. Enzym., **181**, in the press.
28. Goodall, G.J. and Filipowicz, W. (1989) Cell, **58**, 473–483.
29. Bogenhagen, D.F. and Brown, D.D. (1981) Cell, **24**, 261–270.
30. Kiss, T., Antal, M. and Solymosy, F. (1987) Nucleic Acids Res. **15**, 543–560.
31. Brow, D.A. and Guthrie, C. (1988) Nature, **334**, 213–218.
32. Szukálek, A., Kiss, T. and Solymosy, F. (1990) Nucleic Acids Res., **18**, 1295
33. Mühlbach, H.-P. and Sänger, H.L. (1979) Nature, **278**, 185–188.
34. Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) Proc. Natl. Acad. Sci. USA, **83**, 8575–8579.
35. Reddy, R., Henning, D., Das, G., Harless, M. and Wright, D. (1987) J. Biol. Chem. **262**, 75–81.
36. Krol, A., Carbon, P., Ebel, J.-P. and Appel, B. (1987) Nucleic Acids Res., **15**, 2463–2478.
37. Moenne, A., Camier, S., Anderson, G., Margottin, F., Beggs, J. and Sentenac, A. (1990) EMBO J., **9**, 271–277.
38. Tyler, B.M. (1987) J. Mol. Biol., **196**, 801–811.
39. Marshallsay, C., Kiss, T. and Filipowicz, W. (1990) Nucleic Acids Res., **18**, 3459–3466.
40. Kunkel, G.R. and Pederson, T. (1989) Nucleic Acids Res., **17**, 7371–7379.
41. Parry, H.D., Scherly, D. and Mattaj, I.W. (1989) Trends Biochem. Sci., **14**, 15–19.
42. Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. (1985) Ann. Rev. Gen., **19**, 355–387.
43. Stefano, J.E. and Gralla, J.D. (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 1069–1072.
44. Mandecki, W. and Reznikoff, W.S. (1982) Nucleic Acids Res., **10**, 903–912.
45. Murphy, S., Moorefield, B. and Pieler, T. (1989) Trends Genet., **5**, 122–126.
46. Reiter, W.-D., Palm, P. and Zillig, W. (1988) Nucleic Acids Res., **16**, 1–19.
47. Thomm, M. and Wich, G. (1988) Nucleic Acids Res., **16**, 151–163.
48. Mémet, S., Saurin, W. and Sentenac, A. (1988) J. Biol. Chem., **263**, 10048–10051.
49. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R.A. and Zillig, W. (1989) Proc. Natl. Acad. Sci. U.S.A., **86** 4569–4573.