Plant Physiol. (1993) 103: 525-533

Molecular Characterization of PAB2, a Member of the Multigene Family Coding for Poly(A)-Binding Proteins in *Arabidopsis thaliana*[1]

Pierre Hilson, Kathleen L. Carroll, and Patrick H. Masson*

Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706 (P.H., K.L.C., P.H.M.); and Unit of Molecular Biology and Animal Physiology, Faculty of Agronomical Sciences, Gembloux 5030, Belgium (P.H.)

The poly(A) tail of eukaryotic mRNAs associates with poly(A)-binding (PAB) proteins whose role in mRNA translation and stability is being intensively investigated. Very little is known about the structure and function of the PAB genes in plants. We have cloned multiple PAB-related sequences from *Arabidopsis thaliana*. Results suggest that PAB proteins are encoded by a multigene family. One member of this family (PAB2) is expressed in root and shoot tissues. The complete nucleotide sequence of PAB2 was determined. Study of the predicted PAB2 protein reveals a similarity in structure among vertebrate, insect, yeast, and plant PAB proteins. All contain two highly conserved domains: an amino-terminal sequence formed by four RNA recognition motifs and an uncharacterized carboxyl-terminal region of 69 to 71 amino acids. Possible roles for the carboxyl-terminal conserved domain are discussed in view of recently published data concerning the structure and function of PAB proteins.

The poly(A) tail of eukaryotic messenger RNAs is bound by poly(A)-binding (PAB) proteins. This association between PAB and poly(A) is an important determinant in the control of mRNA turnover and translation efficiency (Bernstein and Ross, 1989; Jackson and Standart, 1990; Sachs, 1990).

PAB proteins are involved in the control of mRNA stability. They seem to protect polyadenylated mRNAs from nucleolytic degradation. Indeed, PAB depletion results in a decrease of mRNA stability in cell-free systems (Bernstein et al., 1989). Furthermore, various short-lived transcripts are quickly degraded after losing their poly(A) tail (Peltz and Jacobson, 1992). On the other hand, the yeast PAB protein is required for the shortening of long poly(A) tails by the poly(A) nuclease (Sachs and Davis, 1989; Sachs and Deardoff, 1992).

Several lines of evidence suggest that the PAB protein is involved in the control of mRNA translation. First, polyadenylated mRNAs are more efficiently translated than their poly(A)$^-$ counterparts provided PAB proteins are present (Munroe and Jacobson, 1990). Second, polyadenylated mRNAs are preferentially incorporated into heavy polysomes (Munroe and Jacobson, 1990). Third, seven extragenic suppressors of a temperature-sensitive mutation in the *Saccharomyces cerevisiae* PAB locus all affect the accumulation of the 60 S ribosomal unit (Sachs and Davis, 1989, 1990). Finally, the transient depletion of PAB or PAN in yeast results in the lengthening of poly(A) tails, together with a decrease in the number of ribosomes associated with mRNAs (Sachs and Davis, 1989; Sachs and Deardoff, 1992). This last result suggests that poly(A) tail shortening and translation initiation are interrelated processes.

A gene encoding PAB was first characterized in *S. cerevisiae* where it is essential for cell viability (Adam et al., 1986; Sachs et al., 1986). Homologs of the *S. cerevisiae* PAB gene have been identified in human (Grange et al., 1987), *Xenopus laevis* (Nietfeld et al., 1990; Zelus et al., 1989), *Drosophila melanogaster* (Lefrère et al., 1990), and *Schizosaccharomyces pombe* (Burd et al., 1991). Their amino-terminal portion contains four RRMs, which characterize many RNA-binding proteins other than PAB. They consist of conserved sequences of 80 amino acids carrying two smaller and highly conserved sequences named RNP1 (eight residues) and RNP2 (six residues), which are separated by a less-conserved region of about 30 amino acids (Kenan et al., 1991). In this report we describe the molecular cloning and characterization of *Arabidopsis thaliana* genes coding for PAB proteins.

MATERIALS AND METHODS

Plant Materials

For DNA extraction, seeds of *Arabidopsis thaliana* (L.) Heynh. (Columbia ecotype) were germinated and the plantlets were grown for 2 to 3 weeks in flasks containing liquid Murashige Minimal Organics Medium (GIBCO) complemented with Gamborg's vitamins (Sigma), under a 16-h light/8-h dark photoperiod, at 24°C. The same material was

[1] During part of this work, P.H.M. was mandated (chercheur qualifié) by, and P.H. was a fellow (chargé de recherches) of the Fonds National de la Recherche Scientifique of Belgium. P.H. was also supported by a long-term fellowship from the North Atlantic Treaty Organization. This work was partially supported by grant 2.4570.91 from the Fonds de la Recherche Fondamentale Collective of Belgium to P.H.M. and by start-up funds provided by the College of Agriculture and Life Sciences, the Graduate School and the Laboratory of Genetics at the University of Wisconsin, Madison. This is paper number 3365 of the Laboratory of Genetics.

* Corresponding author; fax 1-608-262-2976.

Abbreviations: CTC domain, carboxyl-terminal conserved domain; PAB, poly(A)-binding; PAN, poly(A) nuclease; PCR, polymerase chain reaction; poly(A), polyadenylic acid; RRM, RNA-recognition motif; RNP, ribonucleoprotein conserved sequence; SSC, standard sodium citrate; UTR, untranslated region.

used for root RNA extraction after removal of the upper green tissues. For shoot RNA extraction, seeds of the same ecotype were germinated, and plants were grown for 3 weeks in soil under the same photoperiod and temperature conditions, at 50% RH.

Nucleic Acid Isolation and Hybridization

For Southern blot analyses, genomic DNA was extracted from *A. thaliana* (Columbia ecotype) as described by Dellaporta et al. (1984), digested to completion with appropriate restriction enzymes, size fractionated by electrophoresis in 0.8% agarose gels in TBE buffer (90 mM Tris, 90 mM borate, 2 mM EDTA), transferred to nylon membranes (Hybond-N+; Amersham), and hybridized to specific DNA probes in a solution containing 6× SSC, 0.5% (w/v) SDS, 5× Denhardt's solution, 20 $\mu g\ mL^{-1}$ of autoclaved herring sperm DNA, and 50% (v/v) formamide for 16 h at 42°C. Low-stringency hybridization experiments were performed as described above, except that formamide concentration was decreased to 30% (v/v). All probes were generated by labeling specific DNA fragments with [α-^{32}P]dCTP (Amersham), using the multiprime labeling system (Amersham). After hybridization the membranes were washed twice in 1× SSC, 0.1% (w/v) SDS at room temperature for 20 min and three times in 0.2× SSC, 0.1% (w/v) SDS at 68°C for 20 min.

For northern blot analyses, *A. thaliana* (Columbia ecotype) RNA was extracted as described by Masson et al. (1989). Shoot poly(A)$^+$ RNA was purified as described by Sambrook et al. (1989). Total and poly(A)$^+$ RNAs were size fractionated by electrophoresis in a Mops-formaldehyde 1.2% agarose gel and transferred to nylon membranes. Hybridization to DNA probes and washes of the RNA gel blots were the same as for DNA gel blots.

Isolation of Clones Coding for an RRM

The genomic clones described in this study and coding for an RRM were obtained by PCR amplification of *A. thaliana* genomic DNA (Columbia ecotype) using a pair of degenerate oligonucleotides corresponding to RNP1 [RPP1–4: 5′-CCG GATCCCTITA(CT)GTIAAGAA(CT)CT-3′] and RNP2 [RPP2–1: 5′-TTGAATTC(ACGT)AC(AG)AA(ACGT)(CG) (CAG)(AT)A(ACGT) CC(CT)(CT)T-3′] and Taq DNA polymerase (Promega). This pair of oligonucleotides was one within a set of four pairs of complementary degenerate primers developed for the general purpose of cloning and characterizing *A. thaliana* genes coding for proteins containing one or several RRMs (data not shown). The PCR amplification included 35 cycles, each cycle consisting of a denaturation step at 94°C (30 s), an annealing step at 47°C for the first five cycles and at 55°C for the next 30 cycles (30 s), and an extension step at 72°C (60 s). The PCR reaction products of 120 to 160 bp were size fractionated by electrophoresis in a 12% polyacrylamide gel, purified, and reamplified (Sambrook et al., 1989). The corresponding fragments were again purified after agarose gel electrophoresis, cleaved with *Bam*HI and *Eco*RI, and cloned in a pBluescript vector (Stratagene; Sambrook et al., 1989).

Isolation and Characterization of Genomic Clones

Genomic fragments carrying RRM-related motifs were cloned by screening under low-stringency hybridization conditions (see above) an *A. thaliana* (Columbia ecotype) genomic library constructed in λ-GEM11 (32,000 plaques) with a probe corresponding to one of the cloned PCR-amplified fragments (AF1 probe), as described by Ausubel et al. (1991). Ten positive clones (PAB clones) were isolated, and DNA was extracted from each clone, *Hin*dIII digested, size fractionated by agarose gel electrophoresis, transferred to a Hybond-N+ membrane, and hybridized to the AF1 probe under the same low-stringency hybridization conditions. Three classes of genomic fragments, which differed by the size of the AF1-hybridizing *Hin*dIII fragment, were further characterized: 3.9 kb in PAB2, 1.3 kb in PAB1, and 3.1 kb in PAB4. These three AF1-hybridizing *Hin*dIII fragments (H1, H2, and H3, respectively) were subcloned into pBluescript vectors, and their restriction maps were determined (Sambrook et al., 1989). Smaller AF1-hybridizing restriction fragments were identified for each fragment and subcloned.

A *Bgl*II genomic fragment of 3.0 kb from the PAB2 λ-GEM11 clone was identified by cross-hybridization with the overlapping H1 fragment. This 3.0-kb *Bgl*II fragment was subcloned into a pBluescript vector, and its restriction map was determined.

Isolation of cDNA Clones

To clone cDNAs complementary to the PAB2 transcript, we screened a leaf cDNA library constructed in λZAP (Callis et al., 1990), using AF1-H1 as a probe (Sambrook et al., 1989). No positive plaques were identified after screening more than 400,000 plaques. Therefore, PCR-based strategies were used to clone overlapping pieces of PAB2 cDNAs (Fig. 1). All cDNA amplifications were performed with the Pfu DNA polymerase (Stratagene).

The C1 fragment (Fig. 1) was PCR amplified from the previously screened leaf cDNA library (Callis et al., 1990). Phage particles (5 × 10^6 plaque-forming units) were boiled for 10 min in a solution containing 1% (v/v) Triton X-100, 20 mM Tris-Cl (pH 8.5), and 2 mM EDTA and subjected to

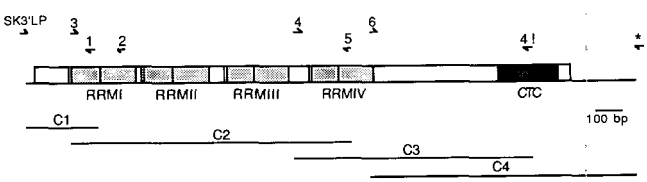

Figure 1. Cloning strategy of the PAB2 cDNA. The structure of the reconstructed cDNA is shown on top with open boxes delineating the open reading frame. The gray boxes represent the RRMs in which RNP1 and RNP2 are represented by vertical lines. The dark box represents the CTC region. The C1 to C4 lines represent the four overlapping cDNA fragments cloned for sequence determination. The arrows show the primers used for PCR amplification: the λZAP-specific primer (SK3′LP), the PAB2-specific primers (HI01–HI06) labeled with their numbers, and the poly(A)-specific primer (dT17-adapter) labeled with a star.

two successive rounds of PCR amplification. The first round included 25 cycles (30 s at 94°C, 60 s at 50°C, and 60 s at 72°C) using as primers SK3'LP (λZAP specific; 5'-GTAATAGATCTCACTATAGGGCGAAT-3') and HI02 (PAB2 specific; 5'-CCATCTAGAGGTATGTAATTCAGTTCTTGG-3'). The product of this primary amplification was subjected to a second round of 40 cycles (same as above) using as primers SK3'LP and HI01 (PAB2 specific and upstream of HI02; 5'-AGACTCGAGCGGTGACCACAGTACCCATTTGGC-3'). The resulting C1 fragment was cleaved with the *Xho*I and *Eco*RI restriction enzymes and cloned in a pBluescript vector (Sambrook et al., 1989).

Additional overlapping fragments of the PAB2 cDNA (C2, C3, and C4; Fig. 1) were amplified and cloned using the rapid amplification of cDNA ends procedure described by Frohman (1990) with cDNA templates synthesized from poly(A)$^+$ RNA extracted from 3-week-old flowering plants (see above). The cycle conditions were as described by Frohman (1990), except that the annealing temperature in each cycle was modified as follows: 52°C for C2 and C3 amplifications and 37°C for the C4 amplification. C2 was primed using the oligonucleotides HI03 (5'-TCGAATTCAATGTGACGGAC-3') and HI05 (5'-GCTTCAGCTGCTTCTTCGGGAGTT-3') and cloned as an *Eco*RI-*Pvu*II fragment in a pBluescript vector. C3 was fortuitously amplified with the HI04 primer at both ends (5'-CAGAATTTGAAGGAAGCTGCAGAC-3') because a stretch of nine nucleotides at the 3' end of HI04 was identical with a sequence in the noncoding strand (Fig. 1). C3 was cloned as a *Pst*I fragment in a pBluescript vector. The C4 3'-end fragments were primed with HI06 (5'-CACAGCTGAAGGAAGACAGAAGG-3') and a mixture of dT17-adapter (5'-GACTCGAGTCGACATCGATTTTTTTTTTTT-TTTTT-3') and adapter (5'-GACTCGAGTCGACATCG-3') (Frohman, 1990). C4 fragments were cloned as *Pvu*II-*Cla*I fragments in a pBluescript vector.

Sequence Analysis

All sequencing reactions were performed on double-stranded DNA templates, using the Sequenase, version 2.0, sequencing kit (United States Biochemical). All nucleotides were determined by sequencing both DNA strands and, for cDNA fragments, two independently amplified clones.

For PAB2 sequencing, various subclones were generated from the C2, C3, and C4 cDNA fragments, as well as from the genomic DNA clones using a combination of exonuclease III-exonuclease VII deletions (Yanisch-Perron et al., 1985) and restriction subcloning procedures (Sambrook et al., 1989). The comparison of the PCR-amplified cDNA clone sequences with the corresponding genomic clone sequence resolved ambiguities and located all intron-exon boundaries. The size of the introns was determined by combining sequence data and restriction fragment length analyses.

Sequence analyses were performed with software from DNASTAR (Madison, WI) and Genetics Computer Group (Madison).

Mapping of the PAB2 Transcript 5' Termini

The 5' end of the PAB2 transcript was located on the genomic sequence by exonuclease VII and S1 endonuclease protection assays and by cDNA primer extension analysis. Both protection assays used a 5' end-labeled single-stranded DNA probe generated as described by Sambrook et al. (1989). Briefly, the oligonucleotide HI09 (5'-GTCTGACCCTGAAGTTGAACCTG-3') was 5' labeled by polynucleotide kinase (Boehringer Mannheim) phosphorylation and annealed to the probe template consisting of a denatured plasmid vector carrying the PAB2 *Bgl*II genomic fragment previously described. After primer extension with the Klenow fragment of *Escherichia coli* polymerase I (New England Biolabs), a 1.1-kb probe-template hybrid fragment was generated by *Hin*dIII digestion. The probe was size fractionated by electrophoresis in a 4% polyacrylamide sequencing gel and purified by electroelution. The labeled 5' nucleotide of the probe is complementary to the 29th nucleotide in the PAB2 mRNA-coding region. For all protection assay samples, the heat-denatured probe (2×10^4 cpm) was hybridized to 30 μg of total RNA in 20 μL of a solution containing 80% (v/v) formamide, 40 mM Pipes (pH 6.4), 400 mM NaCl, and 1 mM EDTA, at 30°C for 16 h. The hybridization reactions were treated either with S1 endonuclease (GIBCO BRL) in 280 mM NaCl, 50 mM sodium acetate (pH 4.5), 4.5 mM $ZnSO_4$, and 20 μg mL^{-1} of denatured herring sperm DNA at 30°C for 1 h or with 30 units of exonuclease VII (GIBCO BRL) in 30 mM KCl, 10 mM Tris-Cl (pH 7.8), and 10 mM EDTA, at 37°C for 2 h. After ethanol precipitation, DNA-RNA hybrids were dissolved, denatured, and analyzed by electrophoresis in 4% polyacrylamide sequencing gels.

For primer extension analysis, the 5' end-labeled HI09 oligonucleotide (10^5 cpm) was hybridized to 1 μg of shoot poly(A)$^+$ RNA as described above. The primer:RNA hybrids were recovered by ethanol precipitation. The primer was extended in a 30-μL reaction containing 50 mM Tris-Cl (pH 8.15), 40 mM KCl, 6 mM $MgCl_2$, 1 mM of each deoxyribonucleotide triphosphate, 1 mM DTT, 150 μg mL^{-1} of actinomycin D, and 25 units of AMV reverse transcriptase (Boehringer Mannheim), incubated at 42°C for 1 h. The reaction was treated with DNase-free RNase (Boehringer Mannheim) and phenol extracted. The labeled cDNA was ethanol precipitated and analyzed in a 4% polyacrylamide sequencing gel.

RESULTS

Identification of DNA Fragments Similar to Genes Coding for RRM in *A. thaliana*

Degenerate oligonucleotides corresponding to the highly conserved RNP1 and RNP2 sequences were synthesized and used to PCR amplify genomic sequences coding for RRMs (see "Materials and Methods"). Multiple discrete DNA fragments were reproducibly synthesized in these PCRs (data not shown). Their sizes ranged from 0.14 to 0.90 kb. The smallest fragments had the size expected for DNA fragments carrying RRM sequences delimited by RNP1 and RNP2 (Fig. 2A). Larger fragments could derive from PCR artifacts, from RRM sequences interrupted by introns, or from sequences with repeated RRMs.

The 0.14-kb PCR-amplified fragments were purified and cloned in a pBluescript vector. Two clones were subjected to DNA sequence analysis: AF1 and AF2 (Fig. 2, B and C,

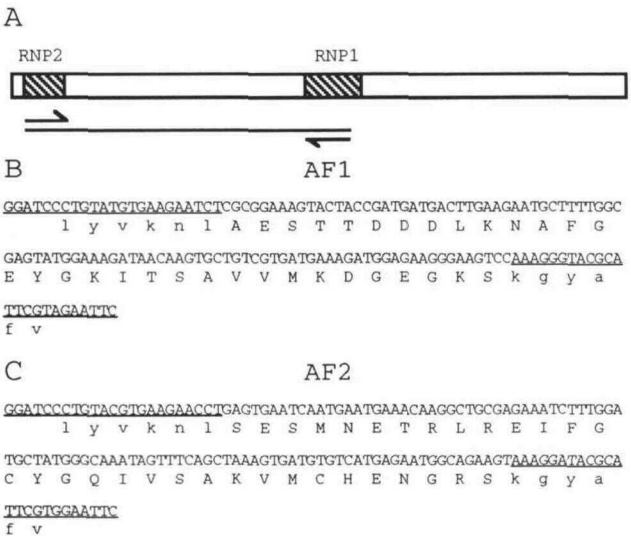

Figure 2. Structure of the AF1 and AF2 genomic clones. A, Schematic representation of the RRM-coding region. The amplified fragment is represented by a line, and the degenerate oligonucleotides are represented as arrows. B and C, DNA and deduced amino acid sequences of AF1 and AF2, respectively. The oligonucleotide sequences are underlined.

respectively). The corresponding deduced amino acid sequences were compared to the protein sequences present in the SwissProt data base (release 20), using the software PROSCAN (default parameters, DNASTAR). The AF1 sequence, between the two RNP motifs, best matches RRM sequences found in the *D. melanogaster* and *S. cerevisiae* PAB proteins, whereas the AF2 sequence best matches the *X. laevis*, *D. melanogaster*, and human PAB proteins (data not shown).

Multiple Sequences Are Related to PAB Genes

The number of AF1- and AF2-related sequences present in the *A. thaliana* genome was investigated by Southern blot analyses. The PCR-amplified AF1 and AF2 fragments were used as probes and hybridized with *Eco*RI- and *Hin*dIII-digested genomic DNA. Under high-stringency hybridization conditions (50% formamide), the AF1 probe hybridized to a unique *Eco*RI fragment of 8.8 kb and to a unique *Hin*dIII fragment of 3.9 kb (H1; Fig. 3 and Table I). Under low-stringency hybridization conditions (30% formamide), the AF1 probe also detected *Eco*RI fragments of 5.0, 4.1, 2.9, and 2.6 kb, as well as *Hin*dIII fragments of 3.1 and 1.3 kb (H3 and H2; Fig. 3 and Table I). On the other hand, the AF2 probe hybridized to an *Eco*RI fragment of 3.1 kb and to a *Hin*dIII fragment of 1.5 kb under high-stringency hybridization conditions. Both AF2-hybridizing fragments differed from the fragments identified by the AF1 probe (Fig. 3 and Table I). No additional AF2-hybridizing sequences were detected under low-stringency hybridization conditions (data not shown).

As a first step in the characterization of genes containing the AF1-hybridizing fragments, an *A. thaliana* genomic library constructed in λ-GEM11 was screened with the AF1 probe under low-stringency hybridization conditions (see "Materials and Methods"). Three positive clones were isolated and further characterized. Each clone contains a different AF1-hybridizing *Hin*dIII fragment whose size corresponds to the size of the H1 (3.9 kb), H2 (1.3 kb), and H3 (3.1 kb) fragments, respectively (Fig. 3 and Table I; data not shown).

To define the relatedness between *Eco*RI and *Hin*dIII AF1-hybridizing fragments, the H1, H2, and H3 fragments were subcloned in pBluescript and tested in Southern blot experiments for their ability to hybridize to *Eco*RI fragments of genomic DNA. The results are summarized in Table I: AF1-H1 hybridizes to an 8.8-kb *Eco*RI fragment, AF1-H2 hybridizes to a 5.0-kb *Eco*RI fragment, and AF1-H3 hybridizes to five small *Eco*RI fragments, including the 2.9- and 2.6-kb AF1-hybridizing fragments.

To identify the genes carried by the cloned H fragments, smaller AF1-hybridizing restriction fragments were identified, subcloned, and partially sequenced (data not shown). Results indicate that AF1-H1 carries a sequence identical with AF1, whereas AF1-H2 and AF1-H3 contain sequences homologous to AF1 (75 and 63% similarity, respectively). It is interesting that the similarity between the three sequences and PAB genes of other organisms extends beyond the AF1-related RRM domain: the region analyzed in AF1-H2 contains a sequence that potentially codes for a peptide homologous to the region extending from RRM I to RRM III in the PAB proteins, whereas the region analyzed in AF1-H3 contains a PAB-related sequence characterized by two rearrangements

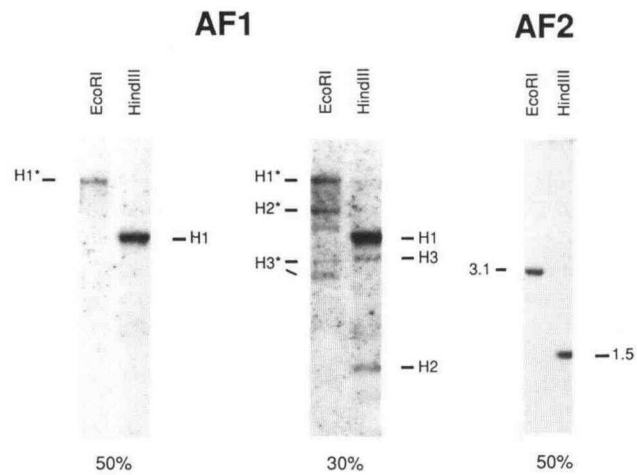

Figure 3. Southern hybridization with *A. thaliana* genomic DNA. The genomic fragments AF1 (A) and AF2 (B) were labeled and hybridized to DNA gel blots. The formamide content (percentage) of the hybridization buffers is indicated at the bottom of the autoradiograms. The AF1-specific bands corresponding to the *Hin*dIII restriction fragments of the PAB2, PAB1, and PAB4 genomic clones are marked H1, H2, and H3, respectively. AF1-specific *Eco*RI restriction fragments also hybridizing to one of the cloned *Hin*dIII fragments (H1, H2, or H3 probes) are positioned and labeled with the name of the homologous H probe and marked with an asterisk. The sizes of the AF1-specific bands are given in Table I, and the sizes of the AF2-specific bands are indicated in kb.

Table I. *EcoRI and HindIII restriction fragments detected by PAB-related probes in Southern blot hybridization experiments with A. thaliana genomic DNA*

The probes are identified at the top of each column.

Restriction Fragment	AF1 (PCR amplified, 137 bp)	PAB2 (AF1-H1, 3.9 kb)	PAB1 (AF1-H2, 1.3 kb)	PAB4 (AF1-H3, 3.1 kb)	AF2 (PCR amplified, 137 bp)
*Hin*dIII	3.9	3.9			
	3.1[a]			3.1	
				2.7	
	1.3[a]		1.3		
*Eco*RI	8.8	8.8			
	5.0[a]		5.0		
	4.1[a]				
					3.1
	2.9[a]			2.9	
	2.6[a]			2.6	
				2.2	
				1.5	
				1.0	

[a] Fragments detected by the AF1 probe under low-stringency hybridization conditions.

(deletions of 37 and 12 bp) that destroy the PAB-homologous open reading frame in a region extending between RRM II and RRM IV (data not shown). Therefore, AF1-H3 seems to contain a PAB pseudogene. Finally, the region sequenced in AF1-H1, together with the flanking sequence cloned as a *Bgl*II fragment (as described in "Materials and Methods"), potentially codes for the entire amino-terminal part of a PAB protein carrying four highly conserved RRMs (see below). The PAB genes or pseudogenes carried by AF1-H1, AF1-H2, and AF1-H3 will be tentatively named PAB2, PAB1, and PAB4, respectively.

Molecular Characterization of an Expressed PAB Gene

Expression of the PAB-related sequences was analyzed by northern blot analysis of total and poly(A)$^+$ RNA extracted from *A. thaliana* flowering shoots, using PAB1-, PAB2-, and AF2-specific probes. As shown in Figure 4, the PAB2-specific probe detected a 2.5-kb transcript. A transcript of the same size was detected in total RNAs extracted from root tissue (data not shown). No transcripts were detected with the PAB1- and AF2-specific probes (data not shown). PAB4 expression was not tested because of its probable pseudogene identity.

The sequence of the PAB2 transcript was determined by cloning overlapping pieces of cDNA (Fig. 1) and by mapping its 5′ termini on the genomic sequence. As shown in Figure 5, the PAB2 transcript consists of nine exons whose structure is not correlated with the organization of known or suspected functional domains in PAB proteins (see below). Nuclease S1 protection assays indicate that PAB2 transcripts share the same 5′ end structure in root and shoot tissues (Fig. 6A). Exonuclease VII and nuclease S1 protection assays together with primer extension experiments demonstrate that the 5′ UTR of the PAB2 is not interrupted by introns and that the 5′ termini of the transcript map in a region 329 to 338 nucleotides upstream of the AUG initiation codon in both shoot and root mRNAs (Fig. 6B; underlined in Fig. 5B). Microheterogeneities observed between assays probably resulted from the different activities of the two nucleases and from the high AT content of the sequence surrounding the transcription start sites identified by primer extension (Fig. 6B). Indeed, the smaller products detected in the nuclease S1 protection assay terminated in a region containing five T residues.

The PAB2 5′ UTR is characterized by clustered stretches of up to eight consecutive A residues. The sequence surrounding the initiation codon is similar to the plant translation start consensus (Fig. 5C; Lütcke et al., 1987). Three alternative polyadenylation sites were identified in a 100-bp region at the 3′ end of the PAB2 transcripts (Fig. 5C).

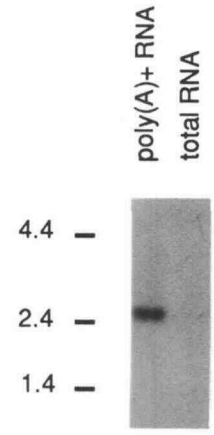

Figure 4. Northern blot analysis of flowering shoot RNAs. A 1.5-kb *Pvu*II-*Sma*I genomic restriction fragment containing the second and third PAB2 exons (see Fig. 5A) was used as a probe.

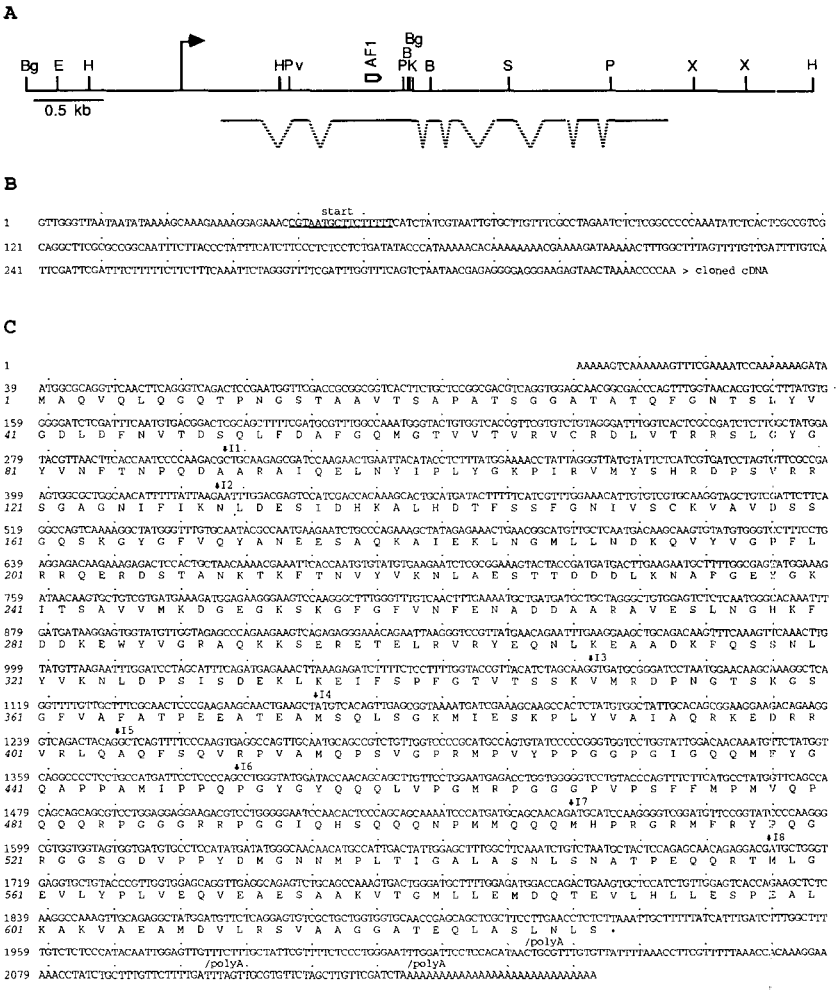

Figure 5. Structure of the PAB2 gene. A, Positioning of PAB2 exons and introns. The upper line shows the restriction map of the PAB2 genomic sequence. The indicated restriction sites are *Bam*HI (B), *Bgl*II (Bg), *Eco*RV (E), *Hin*dIII (H), *Kpn*I (K), *Pst*I (P), *Pvu*II (Pv), *Sma*I (S), and *Xba*I (X). The arrow indicates the transcription start site location. The lower line shows the cloned cDNA sequence interrupted by eight introns. B, Location of the PAB2 transcription start sites in the genomic sequence. The underlined region shows the positions of the major start sites identified by exonuclease VII and endonuclease S1 protection and by primer extension experiments (Fig. 6). C, Sequence of the cloned PAB2 cDNA. Arrows indicate the position of the eight PAB2 introns (I1–I8). Single slashes indicate the poly(A) tail position in the three 3′ end clones analyzed.

Structure of the PAB Protein

The predicted PAB2 protein is 629 amino acid residues long. It has a calculated molecular mass of 68,878 D. To gain more insight into its functional structure, we compared the sequence of the *A. thaliana* PAB2 protein (this report) with the sequences of PAB proteins characterized in *S. cerevisiae* (Sachs et al., 1986), human (Grange et al., 1987), *X. laevis* (Zelus et al., 1989), *D. melanogaster* (Lefrère et al., 1990), and *S. pombe* (Burd et al., 1991). The dot plot matrix comparison between the *A. thaliana* and human PAB proteins shown in Figure 7A is a representative example of the results. Two highly conserved domains can be distinguished. The PAB2 amino-terminal region, which carries the four RRMs, shows a level of identity ranging from 53.4 to 56.2% with equivalent regions in the yeast, human, *X. laevis*, and *D. melanogaster* PAB proteins. Also, the carboxyl-terminal region carries a sequence of 69 to 71 amino acids, which is highly conserved within all of the PAB proteins characterized so far. The latter domain is referred to as the CTC domain (Fig. 7B). The level of identity between CTC domains ranges from 37.1 to 98.5%. This high level of homology between CTC domains is comparable with, and sometimes exceeds, the conservation between RRMs in the same pair of proteins (data not shown).

The region joining the RRMs and the CTC domain is 147 residues long in the *A. thaliana* PAB2 protein. It varies in size (from 92 residues in *S. cerevisiae* up to 190 in *D. melanogaster*) and in sequence between PAB proteins of various organisms. However, despite a lack of sequence similarity, these joining regions share common features: they are rich in Pro and Glu, as well as in Ala and Gly, although the ratio Ala/Gly is variable (data not shown).

DISCUSSION

We have shown that multiple sequences related to PAB genes are present in the *A. thaliana* genome. Most of these PAB-related sequences have not been fully characterized. Therefore, we cannot unambiguously identify them as PAB genes. However, the level of similarity between these sequences and the sequences of known PAB genes is similar to the one found between the previously characterized PAB genes, and it is much higher than with any other gene present in the data base, including other RRM-coding genes (data not shown). Furthermore, Drs. Belostotsky and Meagher (Uni-

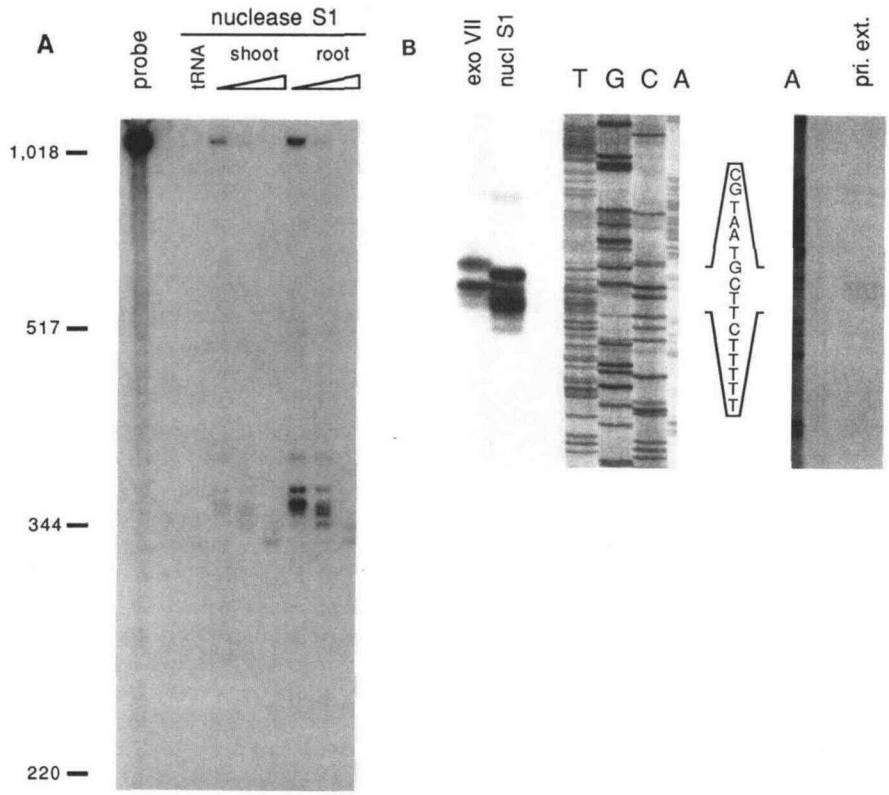

Figure 6. Mapping of the PAB2 transcript 5′ termini. A, Nuclease S1 protection assay. The left lane shows the undigested probe (2×10^3 cpm; described in "Materials and Methods"). For all nuclease-treated samples, the probe (2×10^4 cpm) was hybridized to 30 μg of RNA identified on the top of the figure. The amount of enzyme used was 300 units in the tRNA control and 30, 90, or 300 (left to right) in the shoot and root total RNA samples. The sizes of molecular size markers are indicated at the left in nucleotides. B, Comparison of exonuclease VII (exo VII), nuclease S1 (nucl S1), and primer extension (pri. ext.) experiments. Left, Exonuclease VII and nuclease S1 protection assays. Root total RNA (30 μg) was hybridized to the 5′ end-labeled probe (see "Materials and Methods"). The DNA-RNA hybrids were treated with either 30 units of exonuclease VII or 30 units of nuclease S1. The sequencing reaction was primed with an oligonucleotide identical with the 5′ end of the probe (HI09), annealed to a genomic template. The sequence of the bracketed region, which spans the major signals, is indicated. Right, Result of a primer extension assay using the end-labeled HI09 oligonucleotide as primer on 1 μg of shoot poly(A)$^+$ RNA as template. Both panels are part of the same original autoradiogram. In photographic replica, the right panel was overexposed to allow reproduction of the data. The adenine-sequencing lane was split between the right and left panels to position the signals.

versity of Georgia) have independently cloned and fully characterized an *A. thaliana* flower-specific PAB gene (PAB5) whose sequence differs from all sequences reported in this manuscript. They have partially characterized two other genes, one of which is identical with PAB1 (Belostotsky and Meagher, 1993). Taken together, these data indicate that PAB proteins are encoded by a multigene family in *A. thaliana*.

The significance of the PAB gene multiplicity is uncertain. It could provide more flexibility for spatial and temporal regulation of PAB functions. This hypothesis seems substantiated by the differences observed between the expression patterns of the *A. thaliana* PAB genes characterized so far (this report; Belostotsky and Meagher, 1993). It could also reflect functional differences between closely related proteins (see below).

The PAB2 gene is expressed in both shoot and root tissues. However, no PAB1- or AF2-specific transcripts were identified in these tissues by northern blot analysis. These results suggest that the PAB1 and AF2 genes are not expressed in mature plants. Alternatively, they could be expressed at an undetectable level in these tissues, in a small number of specific cells, or at different developmental stage(s) of the plant. The PAB4 gene carries sequence alterations suggestive of a pseudogene, and, therefore, its expression was not investigated.

The 5′ UTR of the PAB2 transcript contains nine stretches of four to eight adenine nucleotides, most of them clustered. Similar A-rich regions were also found in the 5′ UTR of the *S. cerevisiae* (Sachs et al., 1986), human (Grange et al., 1987), *X. laevis* (Zelus et al., 1989), and *D. melanogaster* (Lefrère et al., 1990) PAB genes. The *S. cerevisiae* and human PAB proteins were shown to bind their own 5′ UTR (Sachs et al., 1986; Berger et al., 1992), suggesting that they might regulate the stability or translational competence of their own transcript. Further experiments are required to test the possible feedback control of PAB2 gene expression by PAB proteins.

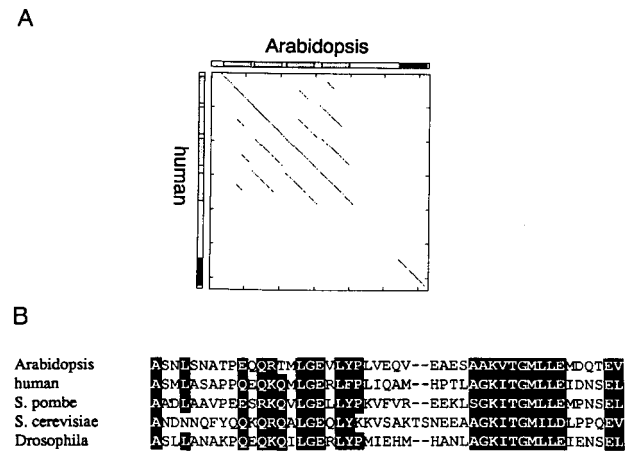

Figure 7. Comparative analysis of PAB proteins. A, Dot plot matrix comparison between the *A. thaliana* and human PAB protein sequences. The amino-terminal ends of both sequences are positioned in the upper left corner of the plot. The gray boxes represent the RRMs and the dark box represents the CTC region. Each diagonal segment indicates regions of 17 residues or more sharing at least 45% identity. The comparison of the amino-terminal portions shows that each RRM is more similar to its counterpart than to the three other RRMs in the homologous protein. B, Alignment of the CTC regions. The boxed sequences indicate conservative changes across the five aligned PAB proteins or identical residues in at least four of them. For each PAB protein, the GenBank accession number of the corresponding gene and the position of the first aligned amino acid (in parentheses) are as follows: *A. thaliana*, L19418 (544); human, Y00345 (544); *S. pombe*, M64603 (560); *S. cerevisiae*, M12780 (494); *D. melanogaster*, L05109 (555). The *D. melanogaster* sequence is part of a revised version of the sequence previously published and now corrected by Lefrère et al. (1990; GenBank accession number L05109). We have verified the corresponding corrections by cloning and sequencing the 3′ end of the coding region in the *D. melanogaster* PAB cDNA (data not shown).

The predicted PAB2 protein is structurally analogous to the PAB proteins characterized in other organisms: the amino-terminal end codes for four conserved RRMs and the carboxyl-terminal end contains a conserved domain of 69 amino acids (CTC domain; Fig. 1). These structural similarities indicate that PAB2 codes for a bona fide PAB protein.

Each individual RRM, including those of PAB2, shows a higher degree of similarity with the equivalent domains in other PAB proteins than with other RRMs in the same protein (Fig. 7A; data not shown). This observation is consistent with previous data showing that specific RRM components exhibit particular binding properties in PAB proteins (Nietfeld et al., 1990; Burd et al., 1991).

The PAB regions joining RRMs and the CTC domain are not conserved but are rich in Pro, Glu, Ala, and Gly. These characteristics suggest that the joining region needs only to be flexible and exposed rather than playing very specific roles in the PAB functions.

The CTC domain of the PAB2 protein is very similar to the same region in all other PAB proteins characterized so far. Because of the wide variety of species considered, the conservation suggests that the CTC domain participates in basic function(s) of the PAB protein. It might be involved in protein-protein interactions, have a function on its own, or act in conjunction with the RRMs to modulate their binding properties. It is interesting that a proteolytic cleavage dissociates the RRM region of PAB from its CTC domain in yeast nuclei (Sachs et al., 1986). Therefore, the CTC domain could be required for some PAB function in the cytoplasm, but it also could be dispensable or even damaging in the nucleus.

The PAB protein controls mRNA degradation by directing the PAN to its substrate (Lowell et al., 1992; Sachs and Deardoff, 1992). This interaction between PAB and PAN has been proposed to occur through a conserved region in the carboxyl-terminal end of PAB (Lowell et al., 1992). CTC could be an excellent candidate for such a conserved region. However, this hypothesis does not take into account earlier reports showing that a truncated PAB protein containing the fourth RRM and the entire carboxyl-terminal region of PAB (including CTC) does not allow poly(A) shortening (Sachs and Deardoff, 1992). If this observation is correct and if the binding property of this truncated protein is unaltered, which is a debated assumption, then one has to conclude that CTC is not sufficient to target PAN to the poly(A) tail of mRNAs (Sachs et al., 1987; Nietfeld et al., 1990; Burd et al., 1991).

From the data discussed above, PAB proteins can be viewed as a bridge between at least two components: the poly(A) tail and PAN. However, it is possible that PAB proteins bind to other RNA sequences or interact with additional protein partners. In this regard, it is interesting to notice that a sequence similar to the CTC domain has been identified in a 100-kD rat protein of unknown function (Müller et al., 1992). Although the specific role of each *A. thaliana* PAB gene is unknown, one can speculate that, if different PAB proteins show distinct association properties, they will control gene expression through different pathways. Experiments studying what effects could result from the depletion of individual members of the PAB multigene family will give clues about their functions.

ACKNOWLEDGMENTS

We are grateful to M. Sepansky for the synthesis of the degenerate oligonucleotides RPP1–4 and RPP2–1, P.M. Hatfield and R.D. Viestra for the gift of an *A. thaliana* leaf cDNA library, A. Laughon for the gift of a *D. melanogaster* embryo cDNA library, A. Burny for providing laboratory space and equipment, T.D. Sullivan for helpful comments and suggestions, D.A. Belostotsky and R.B. Meagher for sharing unpublished data, and T. Stewart for photographic works.

Received April 5, 1993; accepted June 21, 1993.

The GenBank accession number for the sequence reported in this article is (PAB cDNA) L19418.

LITERATURE CITED

Adam SA, Nakagawa T, Swanson MS, Woodruff TK, Dreyfuss G (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol Cell Biol **6:** 2932–2943

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (1991) Current Protocols in Molecular Biology. John Wiley & Sons, New York

Belostotsky DA, Meagher RB (1993) Differential organ-specific expression of three poly(A)-binding protein genes from *Arabidopsis thaliana*. Proc Natl Acad Sci USA **90:** 6686–6690

Berger LC, Bag J, Sells BH (1992) Identification of proteins associating with poly(A)-binding-protein mRNA. Eur J Biochem **204:** 733–743

Bernstein P, Peltz S, Ross J (1989) The poly(A)-poly(A) binding protein complex is a determinant of mRNA stability in vitro. Mol Cell Biol **9:** 659–670

Bernstein P, Ross J (1990) Poly(A), poly(A)-binding protein and the regulation of mRNA stability. Trends Biochem Sci **14:** 373–377

Burd CG, Matunis EL, Dreyfuss G (1991) The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. Mol Cell Biol **11:** 3419–3424

Callis J, Raasch JA, Viestra RD (1990) Ubiquitin extension proteins of *Arabidopsis thaliana*. Structure, localization, and expression of their promoters in transgenic tobacco. J Biol Chem **265:** 12486–12493

Dellaporta SL, Wood J, Hicks JB (1984) Maize DNA miniprep. *In* R Malmberg, J Messing, I Sussex, eds, Molecular Biology of Plants, A Laboratory Course Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 36–37

Frohman MA (1990) RACE: rapid amplification of cDNA ends. *In* M Innis, D Gelfand, J Sninsky, T White, eds, PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, CA, pp 28–38

Grange T, Martins de Sa C, Oddos J, Pictet R (1987) Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif. Nucleic Acids Res **15:** 4771–4787

Jackson RJ, Standart N (1990) Do the poly(A) tail and 3′ untranslated region control mRNA translation? Cell **62:** 15–24

Kenan D, Query CC, Keene JD (1991) RNA recognition: towards identifying determinants of specificity. Trends Biochem Sci **16:** 214–220

Lefrère V, Vincent A, Amalric F (1990) *Drosophila melanogaster* poly(A)-binding protein: cDNA cloning reveals an usually long 3′-untranslated region of the mRNA, also present in other eukaryotic species. Gene **96:** 219–225

Lowell JE, Rudner DZ, Sachs AB (1992) 3′-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev **6:** 2088–2099

Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differ in plants and animals. EMBO J **6:** 43–48

Masson P, Rutherford G, Banks JA, Fedoroff N (1989) Essential large transcripts of the maize *Spm* transposable element are generated by alternative splicing. Cell **58:** 755–765

Müller D, Rehbein M, Baumeister H, Richter D (1992) Molecular characterization of a novel rat protein structurally related to poly(A) binding proteins and the 70K protein of the U1 small nuclear ribonucleoprotein particle (snRNP). Nucleic Acids Res **7:** 1471–1475

Munroe D, Jacobson A (1990) Tales of poly(A): a review. Gene **91:** 151–158

Nietfeld W, Mentzel H, Pieler T (1990) The *Xenopus laevis* poly(A) binding protein is composed of multiple functionally independent RNA binding domains. EMBO J **9:** 3699–3705

Peltz SW, Jacobson A (1992) mRNA stability: in *trans*-it. Curr Opin Cell Biol **4:** 979–983

Sachs A (1990) The role of poly(A) in the translation and stability of mRNA. Curr Opin Cell Biol **2:** 1092–1098

Sachs AB, Bond MW, Kornberg RD (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. Cell **45:** 827–835

Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell **58:** 857–867

Sachs AB, Davis RW (1990) Translation initiation and ribosomal biogenesis: involvement of putative rRNA helicase and RPL46. Science **247:** 1077–1079

Sachs AB, Davis RW, Kornberg RD (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol Cell Biol **7:** 3268–3276

Sachs AB, Deardoff JA (1992) Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell **70:** 961–973

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning, A Laboratory Manual, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene **33:** 103–119

Zelus BD, Giebelhaus DH, Eib DW, Kenner KA, Moon RT (1989) Expression of the poly(A)-binding protein during development of *Xenopus laevis*. Mol Cell Biol **9:** 2756–2760