Plant Physiol. (1994) 106: 303-311

An *Arabidopsis* Chloroplast RNA-Binding Protein Gene Encodes Multiple mRNAs with Different 5′ Ends[1]

Shu-Hua Cheng, Kenneth Cline, and Alice J. DeLisle*

Biochemistry Department, University of Nevada, Reno, Nevada 89557-0014 (S.-H.C., A.J.D.); and Horticultural Sciences Department, University of Florida, Gainesville, Florida 32611-0690 (K.C.)

An *Arabidopsis* cDNA (*Atrbp33*) encoding a nuclear-encoded chloroplast RNA-binding protein (RBP) has been isolated (A.J. DeLisle [1993] Plant Physiol 102: 313-314). ATRBP33 shares global structural homology with all known chloroplast RBPs: a chloroplast transit peptide in the amino terminus, followed by a unique acidic domain and a tandem pair of ribonucleoprotein consensus sequence-type RNA-binding domains in the carboxyl end. In vitro translation products of *Atrbp33* were found to be imported into chloroplasts, suggesting that ATRBP33 is localized in chloroplasts. The expression of *Atrbp33* was higher in chloroplast-containing organs than in nonchloroplast-containing organs. Furthermore, *Atrbp33* was expressed in a light-dependent manner. These features are consistent with its postulated role in posttranscriptional control of chloroplast genes. Northern analyses and RNase protection assays showed that as many as nine messages are encoded by the single *Atrbp33* gene. Sequence analysis of the cDNAs indicated that some of the transcripts have truncated 5′ ends. Most interestingly, the multiple mRNAs potentially encode different polypeptides, one of which lacks a chloroplast transit peptide and acidic domain and contains only one intact RNA-binding domain. Unlike the chloroplast-localized ATRBP33, the truncated polypeptide may function in other cellular compartments.

Chloroplast gene expression requires coordination between plastid and nuclear genomes, because several of the structural, functional, and regulatory components of the chloroplast are encoded in the nucleus (for reviews, see Mullet, 1988; Gruissem, 1989; Taylor, 1989; Rochaix, 1992). Transcription has been demonstrated to be an important regulatory component for chloroplast genes. However, discrepancies between transcription rates and accumulation of messages during the development of chloroplasts indicate that posttranscriptional controls, including RNA processing and stability, play major roles in chloroplast gene regulation (e.g. Deng et al., 1987; Mullet and Klein, 1987; Deng and Gruissem, 1988; Salvador et al., 1993a, 1993b).

It has been shown that the inverted repeat sequences (capable of forming stem-loop structures) in the 3′ untranslated region of spinach chloroplast mRNAs function as RNA processing signals in vitro (Stern and Gruissem, 1987) and stabilize upstream RNA segments in vitro (Stern et al., 1989; Adams and Stern, 1990) and in vivo (Stern et al., 1991). However, the predicted free energies of the stem-loop structure do not correlate with the stability of the cognate mRNAs in vitro (Stern and Gruissem, 1989; Stern et al., 1989). Also, the inverted repeat sequences alone cannot totally account for the developmental, organ-specific, and light-dependent changes in the accumulation and stability of specific chloroplast mRNAs. It was hypothesized that some protein factors may function to process and stabilize plastid mRNAs (Stern and Gruissem, 1989; Stern et al., 1989).

In *Chlamydomonas reinhardtii*, the 5′ ends of mRNAs are thought to be important for message stability (Salvador et al., 1993a); proteins that bind in vitro to the 5′ region of *psbA* mRNA (encoding the thylakoid D1 protein of PSII) have been characterized (Danon and Mayfield, 1991). In addition, many of the factors that contribute to posttranscriptional control of chloroplast genes in *C. reinhardtii* are nuclear encoded (Rochaix, 1992). In several mutants of *C. reinhardtii*, chloroplast messages are transcribed at normal rates compared to wild type, but specific mRNAs are selectively destabilized (Kuchka et al., 1989; Sieburth et al., 1991; Monod et al., 1992). It appears that nuclear gene products in the wild type, which are absent in the mutants, stabilize specific mRNAs.

Nuclear mutants affecting chloroplast gene expression have also been isolated from higher plants. One high-Chl-fluorescence mutant of maize, *hcf7*, is apparently defective in chloroplast rRNA processing, because 16S rRNA precursors accumulate (Barkan, 1993). In barley, nuclear genes have been identified that affect the stability of chloroplast *psbD* messages (Gamble and Mullet, 1989).

In efforts to isolate the protein factors that are involved in the posttranscriptional regulation of chloroplast genes, several groups have purified chloroplast RBPs and isolated the corresponding genes. In spinach, a nuclear-encoded RBP binds to chloroplast mRNA precursors and mature messages at 3′ inverted repeat sequences and is apparently involved in processing the 3′ end of the mRNAs in vitro (Schuster and Gruissem, 1991). Five tobacco nuclear-encoded chloroplast RBPs have also been isolated based on their ability to bind single-stranded DNA (Li and Sugiura, 1990; Ye et al., 1991). Recently, several additional nuclear-encoded chloroplast RBP

[1] This work was supported by U.S. Department of Agriculture-Cooperative State Research Service grant 91-37301-6289 and by funding from the Nevada Agricultural Experiment Station to A.J.D. and by National Institutes of Health grant 1 R01 GM46951 to K.C.

* Corresponding author; fax 1-702-784-1650.

Abbreviations: nt, nucleotide; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RBP, RNA-binding protein; RPA, RNase protection assay; RNP-CS, ribonucleoprotein consensus sequence.

genes were isolated from maize (Cook and Walker, 1992), *Nicotiana plumbaginifolia* (Mieszczak, et al., 1992), and *Arabidopsis* (Bar-Zvi et al., 1992; Didier and Klee, 1992; DeLisle, 1993). All of these chloroplast RBPs share the same domain organization. The amino-terminal chloroplast transit peptide is followed by an acidic domain. The carboxyl-terminal region consists of two RNP-CS-type RNA-binding domains (Bandziulis et al., 1989). It has been demonstrated that the RNP-CS domains are the functional units for RNA-binding activity (Nietfeld et al., 1990; Scherly et al., 1990). In many eukaryotes, the RNP-CS-type RBPs are involved in splicing, polyadenylation, and ribosome biogenesis (Bandziulis et al., 1989; Mattaj, 1989). However, except for the spinach protein (Schuster and Gruissem, 1991), the functions of plant chloroplast RBPs are unknown. It has been postulated that these RBPs are involved in the posttranscriptional processing of chloroplast pre-mRNAs (Li and Sugiura, 1990).

We previously isolated a RBP cDNA from *Arabidopsis* by screening an expression library for single-stranded DNA-binding proteins (DeLisle, 1993). We reported that the cDNA (*Atrbp33*; previously named *Atrbp31*, see DeLisle, 1993) encodes a polypeptide that has the domain organization typically found in all known chloroplast RBPs (Fig. 1A) (Li and Sugiura, 1990; Schuster and Gruissem, 1991; Ye et al., 1991; Bar-Zvi et al., 1992; Cook and Walker, 1992; Didier and Klee, 1992; Mieszczak et al., 1992). The same *Arabidopsis* clones were isolated in other laboratories (Bar-Zvi et al., 1992; Didier and Klee, 1992). The composition of the first 82 amino acids is characteristic of a chloroplast transit peptide: enriched in hydroxylated amino acids (31% Ser and 7% Thr) and basic amino acids (predicted isoelectric point = 12.5). The acidic domain contains the 5-amino acid motif Ser-Glu-Gly-Asp-X repeated six times. Each of the RNA-binding domains consists of approximately 90 amino acids and contains RNP-CS 1 and RNP-CS 2 (Arg-Gly-Phe-Gly-Phe-Val-Thr-Met and Leu/Val-Phe/Tyr-Val-Gly-Asn, respectively). Although there is little sequence homology upstream of the RNA-binding domains, ATRBP33 shares 37 to 76% identity with the RNA-binding domains of chloroplast RBPs in other plants (Li and Sugiura, 1990; Schuster and Gruissem, 1991; Cook and Walker, 1992).

To avoid confusion due to heterogeneity in the *Atrbp33* mRNA 5′ ends (see "Results"), base positions are designated in this report relative to the first ATG start codon (Fig. 1A) (DeLisle, 1993). There are three ATG initiation codons in the same reading frame, at positions +1, +43, and +502 (Fig. 1A). All three ATG codons are in good consensus sequences for translation initiation (Kozak, 1986). The longest ORF encodes a 329-amino acid polypeptide with a calculated molecular mass of 35.8 kD. The second ORF encodes a

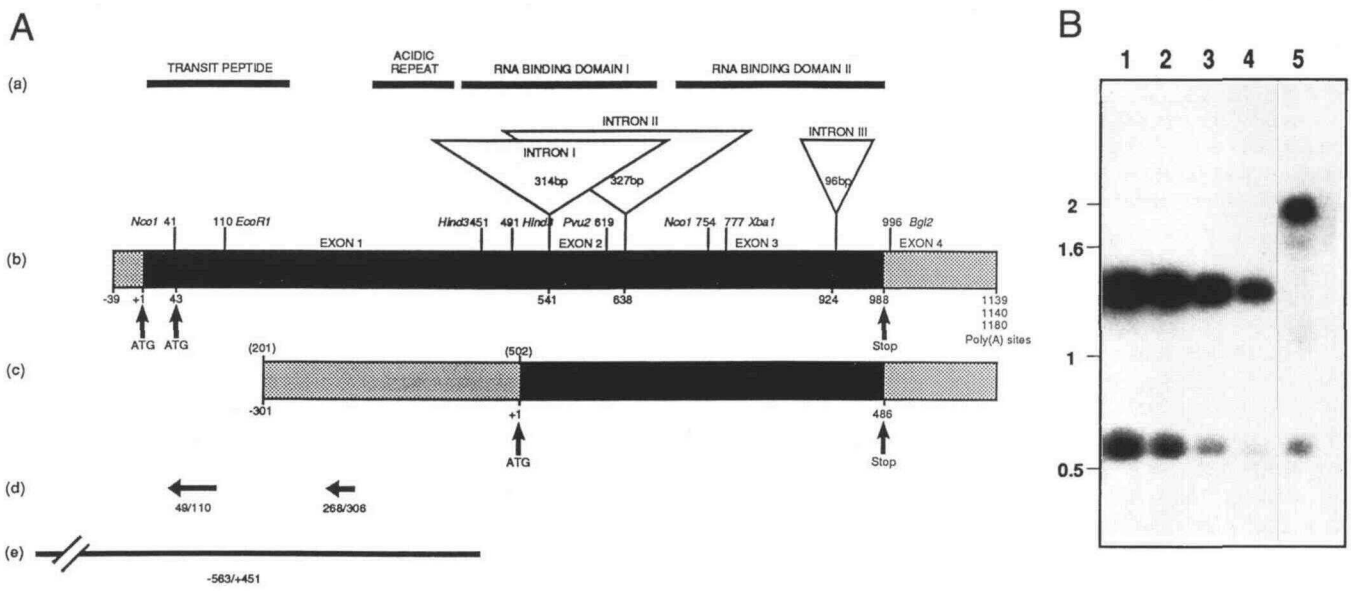

Figure 1. Genomic organization of *Atrbp33*. A, Schematic representation of *Atrbp33* genomic organization and cDNAs, and corresponding protein domains. The GenBank accession number for the *Atrbp33* genomic clone is U08467. The cDNA was previously reported (DeLisle, 1992). Diagrams are shown of cDNAs derived from the two most abundant *Atrbp33* mRNAs from leaves. The potential ORFs are shown in black and the 5′ and 3′ untranslated regions are in gray. Vertical arrows mark start and stop codons. a, Narrow horizontal bars show locations of four structural domains of the encoded proteins. b, A diagram of the cDNA derived from the 1200-nt *Atrbp33* mRNA. Numbers indicate base positions beginning with the first ATG as +1. Locations of selected restriction sites, exons, introns, and known polyadenylation sites are shown. c, A diagram of the cDNA derived from the 1000-nt *Atrbp33* mRNA. Numbers in parentheses are base positions corresponding to those in b. Base positions are numbered using the ATG beginning in the RNA-binding domain I. d and e, Positions of probes used in primer extension analysis and RPA, respectively. B, Copy number reconstruction. Lanes 1 to 4, Three, two, one, and one-half genomic equivalents, respectively, of *Eco*RI-digested plasmid containing the 1200-bp *Atrbp33* cDNA. Lane 5, Five micrograms of *Arabidopsis* genomic DNA digested with *Eco*RI. DNA was size fractionated by agarose gel electrophoresis, blotted to nylon membrane, and probed with a ^{32}P-labeled *Atrbp33* cDNA fragment (see "Materials and Methods"). Migration of molecular mass standards is indicated on the left in kb.

polypeptide of 314 amino acids, and its calculated molecular mass is 34.4 kD. These two polypeptides are identical except that one is 14 amino acids longer than the other at the amino terminus. The third ORF encodes a polypeptide of 167 amino acids with a predicted molecular mass of 18.3 kD. Unlike the longer polypeptides, which consist of all four domains typically found in chloroplast RBPs, the truncated polypeptide contains no transit peptide or acidic domain and has only one intact RNA-binding domain (domain II)

In this report, we present the characterization and expression of the nuclear-encoded chloroplast RBP gene *Atrbp33* from *Arabidopsis thaliana* (DeLisle, 1993). The expression of *Atrbp33* was organ preferential, developmentally regulated, and light dependent. Most interestingly, multiple messages were encoded by *Atrbp33* and one protein encoded by these messages lacked a chloroplast transit peptide. This suggests that the truncated protein may function in cellular compartments other than chloroplasts.

MATERIALS AND METHODS

Cell Cultures and Plant Growth

Arabidopsis thaliana L. (Heyn) Columbia suspension cultures were established and maintained as previously described (Ferl and Laughner, 1989). Leaf materials were obtained from plants grown at 20°C under 10-h/14-h light/dark cycles with a light intensity of 50 to 65 μE m^{-2} s^{-1}. Plants were grown either in potting soil or in 0.6% (w/v) agar containing Murashige and Skoog medium (Murashige and Skoog, 1963) and 3% (w/v) Suc. Roots were isolated from plants grown in agar. Inflorescences were harvested from plants grown in soil with 16-h light/10-h dark photoperiods. Tissues were harvested, immediately frozen in liquid nitrogen, and stored at −70°C until use.

Genomic Library Screening and Sequence Analysis

The *Atrbp33* cDNA (DeLisle, 1993) was used to screen an *Arabidopsis* genomic library (Clontech, Palo Alto, CA) by hybridization (Sambrook et al., 1989). A 5-kb *Sac*I fragment from one of the positive clones was isolated and subcloned into pBluescript KS(−) (Stratagene). DNA sequencing was performed on double-stranded templates by the dideoxy chain-termination procedure (Sanger et al., 1977) using Sequenase 2 (United States Biochemical) and following the manufacturer's protocol. DNA sequences were analyzed using the MacVector program (International Biotechnologies, Inc., New Haven, CT).

Chloroplast Import Assays

The clone containing the 1.2-kb cDNA insert (as shown in Fig. 1A) was transcribed with T3 RNA polymerase in the presence of diguanosine triphosphate and was translated in a wheat germ system with [^{3}H]Leu as label (Cline et al., 1989). Translation was terminated by transfer to an ice bath and dilution to 30 mM unlabeled Leu in import buffer (50 mM Hepes/KOH, pH 8, 0.33 M sorbitol). Intact chloroplasts were isolated from 9- to 10-d-old pea (*Pisum sativum* L. cv Laxton's Progress 9) seedlings (Cline, 1986). The plastids were suspended in import buffer and kept on ice until use. Import into intact chloroplasts was carried out in microcentrifuge tubes for 10 min in a 25°C water bath illuminated with approximately 70 μE m^{-2} s^{-1} (Cline et al., 1989). Assays were terminated by transfer to 0°C. Chloroplasts were then repurified on a Percoll cushion or were first treated with the protease thermolysin and then repurified. For subfractionation, recovered chloroplasts were osmotically lysed with 10 mM Hepes/KOH, pH 8, 10 mM $MgCl_2$ and the soluble fraction was separated from the membrane fraction by centrifugation at 3,200*g* for 8 min. The supernatant was further centrifuged at 40,000*g*. Membranes were either analyzed directly, washed with import buffer, and extracted with 0.1 M NaOH, or treated with thermolysin before analysis. Samples recovered from these assays were subjected to SDS-PAGE and fluorography (Cline, 1986).

RNA Isolation and Northern Hybridization

Total RNA from various tissues was extracted as described (Finkelstein et al., 1985). Leaf poly(A$^+$) RNA was isolated using oligo(dT) cellulose chromatography (type 3, Collaborative Research, Inc., Bedford, MA), size fractionated by electrophoresis in a Mops-formaldehyde 0.8% agarose gel, and transferred to a nylon membrane (Schleicher and Schuell) (Sambrook et al., 1989). A 1-kb cDNA fragment (+110 to +1139, Fig. 1A) was labeled by random priming and was used as a probe for hybridization. The membrane was hybridized in 4× SSC (1× SSC equals 150 mM NaCl and 15 mM sodium citrate, pH 7), 0.1% (w/v) SDS, and 0.25% (w/v) nonfat dry milk at 50°C. Following 12 h of hybridization, the membrane was washed at room temperature in 2× SSC, 0.1% SDS and then in 0.2× SSC and 0.1% SDS at 65°C.

Genomic DNA Isolation and Southern Hybridization

Arabidopsis genomic DNA was isolated from leaves as described (Murray and Thompson, 1980). Five micrograms of genomic DNA were digested with *Eco*RI. Copy number reconstruction experiments were performed with three, two, one, and one-half genomic equivalents of *Eco*RI-digested plasmid containing *Atrbp33* cDNA. DNAs were size fractionated by electrophoresis in 0.8% agarose gels in 90 mM Tris, 90 mM borate, 2 mM EDTA, pH 8, and blotted to a nylon membrane (Sambrook et al., 1989). DNA probes were the same as described for the northern hybridization. The filter was hybridized in 50% (v/v) formamide, 4× SSC, 5× Denhardt's solution, 50 mM sodium phosphate, pH 7, 0.2% SDS, and 250 μg mL^{-1} denatured salmon sperm DNA at 50°C. The washing conditions were the same as described for the northern hybridization.

RPA

Genomic DNA fragments from *Atrbp33* (−524 to +490 relative to the first ATG, Fig. 1A), *AAc1* (actin gene, +1148 to +1830; Nairn et al., 1988), and *Cab140* (Chl *a/b* binding protein gene, +56 to +854; Leutwiler et al., 1986) were cloned into pBluescript KS(−). These plasmid DNAs were linearized

with appropriate restriction enzymes. Radioactive antisense riboprobes were synthesized in vitro using T3 or T7 RNA polymerase (Promega) and purified by 5% (w/v) polyacrylamide/8 M urea gels. RPAs were carried out as described (Hod, 1992) with some modifications. Briefly, 5 μg (unless otherwise indicated) of total RNA were hybridized with the radioactive antisense RNA at 43°C for 16 h in 80% deionized formamide, 100 mM sodium citrate, 300 mM sodium acetate, pH 6.4, and 1 mM EDTA. The hybridized RNA was digested with a mixture of RNase A (10 μg mL^{-1}, United States Biochemical) and RNase T1 (100 units mL^{-1}, United States Biochemical) in 10 mM Tris-HCl, pH 7.5, 300 mM NaCl, and 5 mM EDTA at 37°C for 45 min. The digestion was stopped by adding an equal volume of 4 M guanidine thiocyanate. The undigested RNA was precipitated with an equal volume of isopropanol and the protected fragments were analyzed on 5% PAGE containing 8 M urea. The gels were dried and autoradiographed.

Primer Extension Analysis

Twenty micrograms each of total RNA from leaf and flower were denatured at 70°C for 10 min and hybridized to ^{32}P-labeled primers (positions 49–110 and 268–306, see Fig. 1A) at 49°C for 5 h in 20 μL of 10 mM Tris, pH 7.5, 1 mM EDTA, 250 mM KCl. The primer extension reaction was done by adding 50 μL of reaction mixture containing 14 mM $MgCl_2$, 14 mM Tris-HCl, pH 8.3, 0.28 mM each of dATP, TTP, dCTP, and dGTP, 7 mM DTT, and 7.5 units of avian myeloblastosis virus reverse transcriptase (Life Science, Inc., St. Petersburg, FL) to the hybridization mix and incubating at 42°C for 30 min. The reaction was ethanol precipitated and analyzed on a 6% sequencing gel.

cDNA 5′ End Amplification

RACE was employed to generate the 5′ ends of the cDNAs (Frohman et al., 1988). The amplification was carried out using the 5′ ampliFINDER RACE kit (Clontech) according to the manufacturer's protocol. RNA from leaf and flower was reverse transcribed using the gene-specific primer 5′-TAACCTCGGCGATTTCAACAGTACC-3′. The reverse-transcribed cDNAs were amplified by PCR using an internal nested primer, 5′-AATCCGAAGTCTGGGCAACGAAGG-3′. The amplified products were cloned into pBluescript KS(−) and sequenced as described for the genomic clone.

RESULTS

Gene Structure

The genomic organization of *Atrbp33* was analyzed by digesting genomic DNA with restriction enzymes *Bgl*II, *Eco*RI, and *Hin*dIII and probing with either the entire *Atrbp33* cDNA or fragments specific for the unique acidic domain or the RNA binding domains. In each case, one or two bands were detected (data not shown). Such a simple pattern suggests that *Atrbp33* is encoded by a single locus. Furthermore, a copy number reconstruction experiment indicated that *Atrbp33* is a single-copy gene (Fig. 1B).

A genomic clone containing the entire *Atrbp33* coding region was isolated by hybridization to the cDNA. A 5-kb *Sac*I fragment was subcloned into a Bluescript vector and partially sequenced. The sequenced region (2716 bp) includes 563 bp of 5′ flanking sequence, the structural gene (1881 bp containing introns), and 274 bp of 3′ flanking sequence. There are four exons and three introns of 314, 327, and 96 bp, which interrupt the coding region at positions +541, +638, and +924, respectively (Fig. 1A). The relative positions of the three introns are similar to those in tobacco and maize chloroplast RBP genes (Li et al., 1991, 1993; Cook and Walker, 1992).

In Vitro Chloroplast Import of ATRBP33 Precursor

ATRBP33 must be translocated into chloroplasts to participate in regulation of chloroplast genes. To determine whether the protein is targeted to the chloroplasts, in vitro transcription/translation products of *Atrbp33* were incubated with isolated, intact pea chloroplasts and radiolabeled proteins were analyzed (Cline, 1986; Cline et al., 1989). Two in vitro translation products were found (Fig. 2, lane 1), which probably resulted from translation initiation at the two start codons at positions +1 and +43 (Fig. 1A). The predicted molecular masses of the in vitro translation products were 34.4 and 35.8 kD (Fig. 1A); however, the molecular masses were estimated to be 40 and 42 kD, respectively, by SDS-PAGE. Upon incubating the translation products with intact chloroplasts, a 32-kD band appeared that was recovered with intact chloroplasts (lane 2) and protected from exogenous protease (lane 3). The appearance of this band was dependent on ATP, time, and temperature (the 32-kD form did not accumulate at 0°C; data not shown). Upon chloroplast subfractionation, the 32-kD protein was recovered primarily in the soluble fraction. Although some of the 32-kD protein was recovered with the membrane pellet, it was only periph-

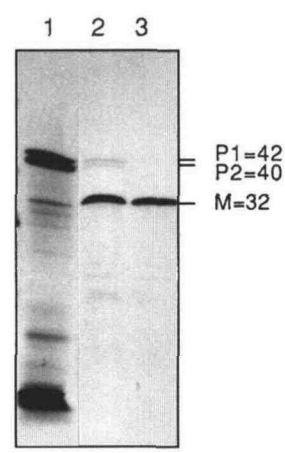

Figure 2. Chloroplast import assay. Chloroplasts were incubated with *Atrbp33* translation products as described in "Materials and Methods." Lane 1, Translation products. Lane 2, Chloroplasts repurified after incubation. Lane 3, Recovered chloroplasts were treated with the protease thermolysin prior to repurification. The apparent molecular masses were calculated by comparison to radiolabeled protein standards (Sigma). A picture of the fluorogram is shown.

erally bound to the membranes. The protein was largely removed upon washing and eliminated by protease treatment of the membranes or extraction with 0.1 M sodium hydroxide (data not shown). These results indicate that the precursor was imported into the chloroplast and proteolytically processed to remove the transit peptide.

Northern Analysis

The expression of *Atrbp33* was analyzed by a northern blot of poly(A^+) RNA isolated from leaves. Under relatively stringent hybridization and washing conditions (see "Materials and Methods"), the *Atrbp33* cDNA hybridized to multiple messages, with 1000- and 1200-nt bands being the most apparent (Fig. 3). Furthermore, labeled transit peptide coding region hybridized only to the 1200-nt mRNA (data not shown), indicating that the sequences at the 5′ ends are different in these messages and that at least the larger mRNA could encode the RBP destined for the chloroplasts.

Mapping 5′ Ends

Since multiple messages with different 5′ ends were produced by *Atrbp33*, an RPA was performed to map the 5′ ends in detail. An antisense riboprobe complementary to the region from 563 bases upstream to 451 bases downstream of the first translation initiation codon was used (Fig. 1A). There were at least nine different RNA fragments protected, ranging in length from 540 to 250 nt (Fig. 4). These RPA fragments were not artifacts of incomplete RNase digestion because the intensity of the fragments increased as the amount of RNA hybridized increased and the intensity of the fragments remained constant with increased RNase concentration (data not shown). According to their sizes, the protected fragments can be divided into two families that correspond to the 1000- and 1200-nt mRNAs detected on the northern blot (Fig. 3). The smaller family contained protected fragments of 250 to 278 nt, and the 250-nt fragment was the most prominent band. The larger family had protected fragments ranging from 427 to 540 nt, with the 490-nt band being most abundant. These results indicate that there are multiple sites of transcription initiation and/or alternative splicing for *Atrbp33*.

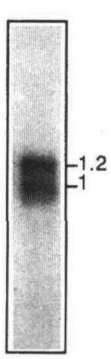

Figure 3. Northern blot analysis. Two micrograms of leaf poly(A^+) RNA were separated on a formaldehyde-agarose gel, blotted to a nylon membrane, and hybridized with a ^{32}P-labeled *Atrbp33* cDNA fragment (see "Materials and Methods"). The sizes of the hybridized bands are indicated on the right in kb.

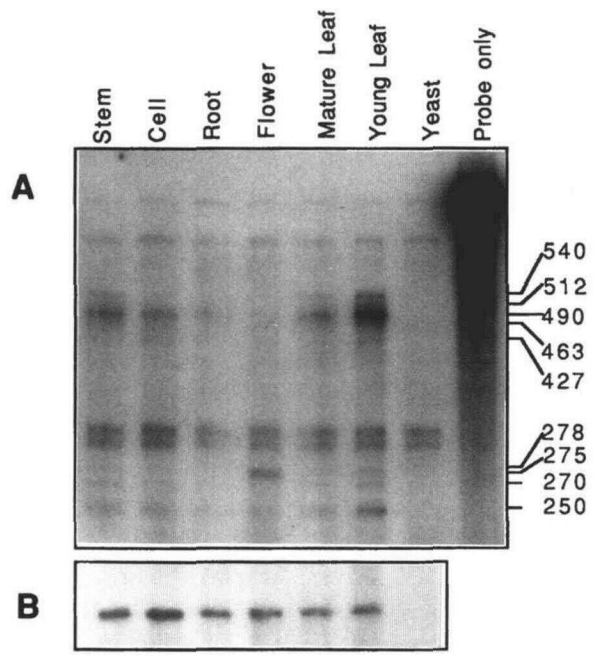

Figure 4. Expression of *Atrbp33* in different organs. A, RNA from the indicated organs was analyzed by RPA using the antisense RNA probe shown in Figure 1A (RPA probe). Cell, Suspension-culture cells. Mature leaf, Leaves of 40-d-old plants. Young leaf, Leaves of 20-d-old plants. Five micrograms of total RNA were analyzed from all organs with the exception of suspension cultures and roots, and 15 μg of total RNA were analyzed from those tissues. Yeast RNA was used as a control. Sizes of protected fragments are indicated on the right in nt. B, RPA of the same RNA as in A, but probed with actin antisense RNA (see "Materials and Methods"). Five micrograms of total RNA were analyzed in each case. The size of the protected fragment is 393 nt.

The various positions of the 5′ ends of the transcripts were verified by primer extension analyses and 5′ RACE PCR and by comparing the results to sequences of cDNAs and genomic clones. The positions of the 5′ ends of the transcripts determined by RPA, 5′ RACE PCR, and primer extension are summarized in Table I. Although not all of the RPA fragments have been confirmed yet by the other two techniques, in general the 5′ ends mapped by the three methods were in close agreement. In particular, two 5′ ends were mapped precisely at the same positions (−89 and −39) by two independent methods. The positions of the majority of the 5′ ends mapped by the different techniques deviated by only a few nts (less than six). The greatest deviation was 10 nts, which occurred with the smallest transcript in the 1200-nt family. Sequence analyses of the 19 independent clones obtained from 5′ RACE PCR so far indicate that some of the messages have truncated 5′ ends and that there is no evidence for differential splicing. The existence of upstream 5′ mini exons, however, remains a possibility.

Despite the various positions of the 5′ ends, all five members of the 1200-nt family encoded a protein that possessed

Table I. *Positions of transcription start sites for Atrbp33*

RNA Species	Method of Detection					
	RPA		5′ RACE PCR		Primer extension	
	Leaf	Flower	Leaf	Flower	Leaf	Flower
1200-nt RNAs	−89	−89		−89		
	−61			−58		
	−39[a]	−39	−33 (3)[b]	−34	−39	
	−22		−18	−18 to − 16 (2)		
	+24		+14			
1000-nt RNAs	+173					
		+176[c]				+177
	+181	+181	+182	+183 (2)		
	+201[d]	+201	+192 to +199 (4)	+191 (2)		+203

[a] The most abundant mRNA in the approximately 1200-nt size range. [b] The numbers in parentheses indicate the frequency of isolation by 5′ RACE PCR. [c] The most abundant mRNA in the flower. [d] The most abundant mRNA in the approximately 1000-nt size range, except in flower.

the same structural framework as other chloroplast RBPs: a chloroplast transit peptide, acidic region, and two RNA-binding domains (Fig. 1A). Interestingly, the second ATG would be used as the translation start codon by the smallest transcript in the 1200-nt family, since this transcript begins at nt +24 and lacks the first ATG. The 5′ ends of the 1000-nt mRNAs begin at about 170 to 200 nt downstream from the first ATG (Fig. 1A). The only ORF found in these transcripts begins at the ATG within the first RNA-binding domain (Fig. 1A). As a result, the predicted protein encoded by the smaller transcripts is in frame with its chloroplast counterpart but is truncated and contains only part of RNA-binding domain I and all of domain II. (Fig. 1A).

Developmental Expression

To examine the expression of each of the *Atrbp33* transcripts in different organs, total RNA extracted from various tissues was analyzed by RPA (Fig. 4A). Note that 3 times as much total RNA from suspension-culture cells and roots was used in RPA due to low abundance of the messages. The expression of the messages was organ preferential, being highest in leaves, followed by stems, inflorescence, suspension-culture cells, and roots. In contrast, the actin gene showed similar expression in the different organs (Fig. 4B). These results demonstrated that the relative abundance of *Atrbp33* mRNAs correlates with the presence of chloroplasts. The 250- and 490-nt protected fragments represented the two predominant transcripts in all vegetative tissues examined, except that only one major *Atrbp33* mRNA was observed in roots (the 490-nt fragment). Interestingly, a unique band of 275 nt was found in the inflorescence RNA sample. This fragment represented the most pronounced transcript in floral tissue; it has not yet been detected by 5′ RACE PCR but was confirmed by primer extension (Table I). Additionally, young leaves also expressed six minor transcripts, some of which were undetectable in other tissues by RPA. Moreover, all of the leaf *Atrbp33* mRNAs were developmentally regulated; young leaves displayed higher expression than mature leaves (Fig. 4A). When using more sensitive techniques such as 5′ RACE PCR, however, some of these minor messages were also found in the inflorescence (Table I).

Light Induction

Many genes for nuclear-encoded, chloroplast-localized proteins, such as the *Cab* gene, which encodes the Chl *a/b* binding protein, are light regulated (Tobin and Silverthorne, 1985). To examine the effect of light on the expression of *Atrbp33*, light-grown plants were subjected to 4 d of darkness and then returned to the light for various lengths of time. Total RNA was isolated from the leaves of the treated plants and analyzed by RPA (Fig. 5A). In the dark the *Atrbp33* messages decreased dramatically to barely detectable levels compared to those in the light. Upon return to the light, increased amounts of *Atrbp33* transcripts were measured after 4 h. The messages were higher after 24 h of continuous light than before the dark treatment. Moreover, all eight *Atrbp33* mRNAs were light inducible in leaves. *Cab* mRNA levels were similarly induced by light, whereas actin mRNA was unaffected (Fig. 5, B and C).

DISCUSSION

In higher plants, chloroplast genes are generally co-transcribed and polycistronic messages are processed into shorter RNA species. The posttranscriptional processing, including 5′ and 3′ end processing and splicing of pre-mRNAs, is thought to be a major regulatory mechanism for chloroplast gene expression (Deng et al., 1987; Mullet and Klein, 1987; Deng and Gruissem, 1988; Salvador et al., 1993a, 1993b). Chloroplast RBPs were first purified from tobacco and are postulated to be involved in the posttranscriptional processing of chloroplast RNAs (Li and Sugiura, 1990). Recently, a spinach chloroplast RBP was shown in vitro to be required for processing of the 3′ ends of chloroplast messages (Schuster and Gruissem, 1991). We have isolated a cDNA clone from *Arabidopsis* encoding a chloroplast RBP (DeLisle, 1993). This protein has the same structural organization as all known chloroplast RBPs and is capable of binding single-stranded DNA. Furthermore, we showed that the ATRBP33

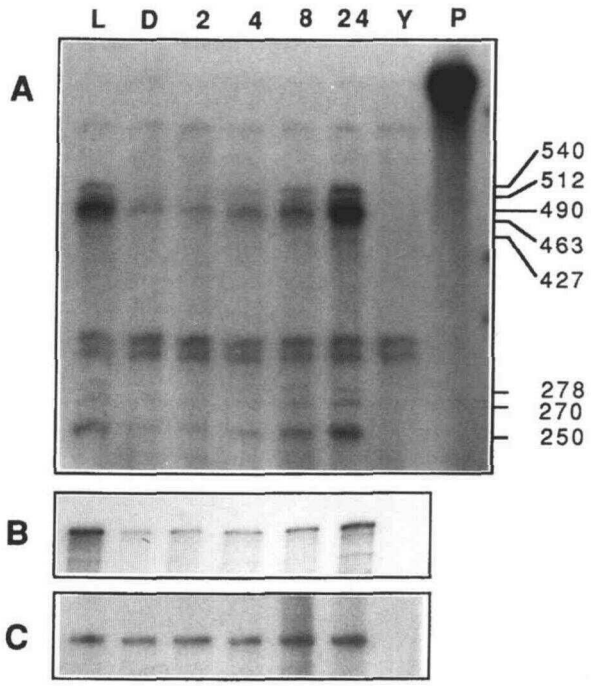

Figure 5. Expression of *Atrbp33* in response to light. A, Five micrograms of total RNA from seedlings were analyzed by RPA, using the *Atrbp33* probe shown in Figure 1A (RPA probe). Plants were grown under a light/dark cycle for 20 d (L), then switched to the dark for 4 d (D). Plants were subsequently returned to the light for 2, 4, 8, and 24 h. Yeast RNA (Y) was used as a control. P indicates the undigested probe. Sizes of protected fragments are indicated on the right in nt. B, RPA of the same RNA as in A, but probed with *cab* antisense RNA (see "Materials and Methods"). The size of the protected fragment is 837 nt. C, RPA of the same RNA as in A, but probed with actin antisense RNA.

protein precursor is imported into isolated chloroplasts and processed to mature size (Fig. 2). The expression of *Atrbp33* mRNAs correlates with the presence of chloroplasts and was regulated by light (Figs. 4 and 5). Based on its structural organization and the present results, we hypothesize that the *Arabidopsis* protein also participates in posttranscriptional events of chloroplast gene expression.

In spite of structural homology between the proteins, the genes encoding chloroplast RBPs from various species exhibit distinct patterns of expression. Three of the five tobacco genes express substantially higher mRNA levels in leaves than in roots (Li and Sugiura, 1990). The expression of the spinach RBP gene (*28rnp*) is highest in leaves, intermediate in cotyledons, and lowest in stems (Schuster and Gruissem, 1991). No *28rnp* messages are detectable in roots. The maize RBP (*NBP*) message accumulates only in leaves, but not in roots, reproductive tissues, or cultured cells (Cook and Walker, 1992). In contrast to the *NBP* gene, *Atrbp33* was expressed in both vegetative and reproductive tissues, and in cultured cells, although its expression was highest in leaves and lowest in roots (Fig. 4). Both the *NBP* and *Atrbp33* genes are developmentally regulated, since higher levels of the transcripts accumulate in young leaves than in mature leaves (Fig. 4A, and Cook and Walker, 1992). Unlike the maize and *Arabidopsis* genes, the expression of the spinach *28rnp* is similar in both young and mature leaves (Schuster and Gruissem, 1991). Furthermore, not all of the chloroplast RBP genes exhibit light-dependent expression. Among the five tobacco genes, only the *cp28* gene is regulated by light (Li and Sugiura, 1990; Ye et al., 1991). There are two spinach *28rnp* messages (1.2 and 1.4 kb) detected by northern analysis (Schuster and Gruissem, 1991), similar to what was observed for *Atrbp33* (Fig. 3). However, only the 1.4-kb spinach transcript is light inducible, whereas all of the *Atrbp33* mRNAs were induced by light (Fig. 5).

It appears that all of these proteins belong to a large family of nuclear-encoded chloroplast RBPs. The different patterns of expression may reflect different functions of each member, a notion supported by the fact that each member contains a highly diverged, unique acidic region. The function of these unique acidic domains is unknown, but they have been suggested to be involved in protein-protein interaction (Preugschat and Wold, 1988; Bar-Zvi et al., 1992). It has been shown that posttranscriptional processing and splicing of pre-mRNAs are complex events and require many ribonucleoproteins (e.g. the heterogeneous nuclear ribonucleoprotein splicing complex) (Bandziulis et al., 1989). Therefore, the functional specification of each chloroplast RBP may be achieved in part by interacting with other processing proteins through the acidic domains.

Didier and Klee (1992) isolated an *Arabidopsis* cDNA, *FMV3bp*, that is probably encoded by the same gene as *Atrbp33*. The nucleic acid sequence of *FMV3bp* is identical to that of *Atrbp33*, except for the first 21 bases, which were derived from a genomic clone (Didier and Klee, 1992). Sequence comparison of *FMV3bp* and the *Atrpb33* genomic clone, however, revealed that the 5′ 21 bases of *FMV3bp* were located in the first intron of the *Atrbp33* genomic clone. This genomic fragment is actually located downstream of where it was placed in *FMV3bp*. Placement of the sequence may be the result of a cloning artifact that occurred during genomic library construction (D. Didier and H. Klee, personal communication). The genomic library used to isolate the 5′ end of *FMV3bp* was constructed using DNA that was partially digested with *Eco*RI. An *Eco*RI site is located at the 3′ end of the 21-base fragment in the intron and may have been ligated with another fragment during library construction. Therefore, *FMV3bp* is probably not an example of an *Atrbp33* mRNA with an alternatively spliced 5′ end.

Despite so much sequence identity, *FMV3bp* hybridized to a single RNA band of about 900 nt on a northern blot, contrasting sharply to the 1000- and 1200-nt mRNAs found in this study (Fig. 3). Furthermore, the highest expression of *FMV3bp* is not in leaves, as observed for *Atrbp33*, but in floral tissue. The reasons for such differences remain to be determined, but the patterns of expression may be influenced by the developmental stage of the tissue used. A notable difference between the two northern blots is that leaf poly(A^+) RNA isolated from reproductive-stage plants was used for hybridization of *FMV3bp*, whereas leaf poly(A^+) RNA isolated from the vegetative stage was used for *Atrbp33*. Perhaps the predominant 1000-nt mRNA that we observed in flowers is also most abundant in other organs of the reproductive plants.

Atrbp33 is a single-copy nuclear gene (Fig. 1B) that encodes as many as nine transcripts with different 5′ ends (Fig. 4A). The heterogeneity of the messages is likely due to multiple transcription start sites. Although processing of the primary transcripts is a formal possibility, there is no precedent for 5′ end processing of eukaryotic mRNAs. In spinach two messages are also encoded by the single *28rnp* gene, but the mechanism by which multiple mRNAs are generated is unknown (Schuster and Gruissem, 1991). The promoter regions of both families of *Atrbp33* transcripts contain no apparent TATA sequences. The absence of TATA sequences and the presence of multiple initiation sites are characteristics frequently identified in housekeeping genes (Dynan, 1986). Genes transcribed from promoters lacking TATA boxes commonly are expressed at relatively low levels, are subjected to little regulation, and show little tissue-specific expression (Dynan, 1986). Multiple transcription start sites were also observed for one of the tobacco chloroplast RBP genes (*cp33*) (Li et al., 1993). The *cp33* message is much less abundant compared with other tobacco RBP messages (e.g. *cp28*) and is not light regulated, although the expression is higher in leaves than in roots (Li and Sugiura, 1990). In contrast, *Atrbp33* was transcribed at moderate levels and was both developmentally regulated and light regulated, and its expression showed organ preference.

In this study we showed that multiple messages are generated by the single *Atrbp33* gene. Sequence analyses of cDNAs indicated that these transcripts potentially encode two distinct proteins (Fig. 1A). The larger messages encode a protein that likely functions as a RBP in chloroplasts. If the 1000-nt mRNAs are translated, the protein product is likely to function in different cellular compartments due to the absence of a chloroplast transit peptide. Different transcription start sites in the tobacco *cp33* gene also result in multiple messages, but they differ only in the 5′ untranslated region and thus encode the same protein (Li et al., 1993). The occurrence of the single *Atrbp33* gene encoding two distinct proteins targeted to different cellular compartments is apparently unique so far among the known chloroplast RBP genes.

Little is known about proteins involved in RNA processing in other organelles in higher plants. Genes encoding proteins with only one RNA-binding domain have been isolated from maize (Ludevid et al., 1992) and recently from tobacco (Hirose et al., 1993). The predicted proteins contain a single RNP-CS-type RNA-binding domain in the amino terminus and a Gly-rich domain in the carboxyl terminus. Proteins synthesized in vitro from the maize and tobacco genes are capable of binding RNAs, indicating that they are RBPs (Ludevid et al., 1992; Hirose et al., 1993). Furthermore, it was demonstrated with the yeast poly(A)-binding protein that one of the four RNA-binding domains is sufficient for cell viability and RNA binding specificity (Sachs et al., 1987). Taken together, these data lend support for the idea that the truncated ATRBP33 protein with only one intact RNA-binding domain may also function as a RBP in *Arabidopsis*. Multiple messages from a single RBP gene are common in other organisms and are produced through a combination of alternative transcription start sites and differential splicing (for review, see Bandziulis et al., 1989; Smith et al., 1989). This is a mechanism by which a single gene can encode multiple proteins with specific functions in RNA processing, thereby obviating the need for a much larger number of individual genes.

All of the *Atrbp33* mRNAs in leaves increased and decreased in parallel in response to light/dark treatments. This suggests that the two distinct proteins encoded by *Atrbp33* have coordinated functions. One likely subcellular location of the truncated ATRBP33 is the cytoplasm, since the protein contains no apparent nuclear localization signals or other signal sequences. If so, the truncated protein may function to stabilize the 1200-nt *Atrbp33* transcripts in the cytosol and thus affect the expression of chloroplast genes in an autoregulatory capacity. We have raised antibodies against the amino terminal region unique to the chloroplast form of ATRBP33 and are raising antibodies against the RNA-binding domain present in both forms of the proteins. Future immunolocalization studies will help to address the functions of the protein products of *Atrbp33*.

ACKNOWLEDGMENTS

We thank JoAne Schwyhart for technical assistance and DNA sequencing. The *Cab140* clone was a generous gift from Dr. Elaine Tobin, and the actin clone was kindly provided by Drs. C. Joseph Nairn and Robert Ferl.

Received February 7, 1994; accepted May 11, 1994.

LITERATURE CITED

Adams CC, Stern DB (1990) Control of messenger RNA stability in chloroplasts by 3′ inverted repeats: effects of stem loop mutations on degradation of *psbA* messenger RNA in vitro. Nucleic Acids Res **18:** 6003–6010

Bandziulis R, Swanson M, Dreyfuss G (1989) RNA-binding proteins as developmental regulators. Genes Dev **3:** 431–437

Barkan A (1993) Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. Plant Cell **5:** 389–402

Bar-Zvi D, Shagan T, Schindler U, Cashmore AR (1992) RNP-T, a ribonucleoprotein from *Arabidopsis thaliana*, contains 2 RNP-80 motifs and a novel acidic repeat arranged in an alpha-helix conformation. Plant Mol Biol **20:** 833–838

Cline K (1986) Import of proteins into chloroplasts: membrane integration of a thylakoid precursor protein reconstituted in chloroplast lysates. J Biol Chem **216:** 14804–14810

Cline K, Fulsom DR, Viitanen PV (1989) An imported thylakoid protein accumulates in the stroma when insertion into thylakoids is inhibited. J Biol Chem **264:** 14225–14232

Cook WB, Walker JC (1992) Identification of a maize nucleic acid-binding protein (NBP) belonging to a family of nuclear-encoded chloroplast proteins. Nucleic Acids Res **20:** 359–364

Danon A, Mayfield SPY (1991) Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins. EMBO J **10:** 3993–4001

DeLisle AJ (1993) RNA-binding protein from *Arabidopsis*. Plant Physiol **102:** 313–314

Deng X-W, Gruissem W (1988) Constitutive transcription and regulation of gene expression in non-photosynthetic plastids of higher plants. EMBO J **7:** 3301–3308

Deng X-W, Stern DB, Tonkyn JC, Gruissem W (1987) Plastid run-on transcription: application to determine the transcriptional regulation of spinach plastid genes. J Biol Chem **262:** 9641–9648

Didier DK, Klee HJ (1992) Identification of an *Arabidopsis* DNA-binding protein with homology to nucleolin. Plant Mol Biol **18:** 977–979

Dynan WS (1986) Promoters for housekeeping genes. Trends Genet **2:** 196–197

Ferl RJ, Laughner BJ (1989) In vivo detection of regulatory factor binding sites of *Arabidopsis thaliana Adh*. Plant Mol Biol **12:** 357–366

Finkelstein R, Tenbarge K, Shumway J, Crouch M (1985) Role of ABA in maturation of rape seed embryos. Plant Physiol **78:** 630–636

Frohman M, Dush M, Martin G (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA **85:** 8998–9002

Gamble PE, Mullet JE (1989) Blue light regulates the accumulation of two *psbD-psbC* transcripts in barley chloroplasts. EMBO J **8:** 2785–2794

Gruissem W (1989) Chloroplast gene expression: how plants turn their plastids on. Cell **56:** 161–170

Hirose T, Sugita M, Sugiura M (1993) cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco—occurrence of tissue-specific alternative splicing. Nucleic Acids Res **21:** 3981–3987

Hod Y (1992) A simplified ribonuclease protection assay. BioTechniques **13:** 852–854

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:** 283–292

Kuchka MR, Goldschmidt-Clermont M, van Dellewijn J, Rochaix J-D (1989) Mutation at the *Chlamydomonas* nuclear *NAC2* locus specifically affects stability of the chloroplast *psbD* transcript encoding polypeptide D2 and PS II. Cell **58:** 869–876

Leutwiler LS, Meyerowitz EM, Tobin EM (1986) Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*. Nucleic Acids Res **14:** 4051–4064

Li Y, Sugiura M (1990) Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs. EMBO J **9:** 3059–3066

Li Y, Ye L, Sugita M, Sugiura M (1991) Tobacco nuclear gene for the 31 kD chloroplast ribonucleoprotein: genomic organization, sequence analysis and expression. Nucleic Acids Res **19:** 2987–2991

Li YQ, Nagayoshi S, Sugita M, Sugiura M (1993) Structure and expression of the tobacco nuclear gene encoding the 33 kDa chloroplast ribonucleoprotein. Mol Gen Genet **239:** 304–309

Ludevid MD, Freire MA, Gomez J, Burd CG, Albericio F, Giralt E, Dreyfuss G, Pages M (1992) RNA binding characteristics of a 16 kDa glycine-rich protein from maize. Plant J **2:** 999–1003

Mattaj I (1989) A binding consensus: RNA-protein interactions in splicing, snRNPs, and sex. Cell **57:** 1–3

Mieszczak M, Klahre U, Levy JH, Goodall GJ, Filipowicz W (1992) Multiple plant RNA binding proteins identified by PCR: expression of cDNAs encoding RNA binding proteins targeted to chloroplasts in *Nicotiana plumbaginifolia*. Mol Gen Genet **234:** 390–400

Monod C, Goldschmidt-Clermont M, Rochaix J-D (1992) Accumulation of chloroplast *psbB* RNA requires a nuclear factor in *Chlamydomonas reinhardtii*. Mol Gen Genet **231:** 449–459

Mullet J (1988) Chloroplast development and gene expression. Annu Rev Plant Physiol Plant Mol Biol **39:** 475–502

Mullet J, Klein R (1987) Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. EMBO J **6:** 1571–1579

Murashige T, Skoog F (1963) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant **15:** 473–497

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res **8:** 4321–4325

Nairn CJ, Winesett L, Ferl RJ (1988) Nucleotide sequence of an actin gene from *Arabidopsis thaliana*. Gene **65:** 247–257

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:** 3699–3706

Preugschat F, Wold B (1988) Isolation and characterization of a *Xenopus laevis* C protein cDNA: structure and expression of a heterogeneous nuclear ribonucleoprotein core protein. Proc Natl Acad Sci USA **85:** 9669–9673

Rochaix J-D (1992) Post-transcriptional steps in the expression of chloroplast genes. Annu Rev Cell Biol **8:** 1–28

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

Salvador ML, Klein U, Bogorad L (1993a) 5′ sequences are important positive and negative determinants of the longevity of *Chlamydomonas* chloroplast gene transcripts. Proc Natl Acad Sci USA **90:** 1556–1560

Salvador ML, Klein U, Bogorad L (1993b) Light-regulated and endogenous fluctuations of chloroplast transcript levels in *Chlamydomonas*—regulation by transcription and RNA degradation. Plant J **3:** 213–219

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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA **74:** 5463–5467

Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B and their cognate RNAs. Nature **345:** 502–506

Schuster G, Gruissem W (1991) Chloroplast mRNA 3′ end processing requires a nuclear-encoded RNA-binding protein. EMBO J **10:** 1493–1502

Sieburth LE, Berry-Lowe S, Schmidt GW (1991) Chloroplast RNA stability of *Chlamydomonas*: rapid degradation of *psbB* and *psbC* transcripts in two nuclear mutants. Plant Cell **3:** 175–189

Smith C, Patton J, Nadal-Ginard B (1989) Alternative splicing in the control of gene expression. Annu Rev Genet **23:** 527–577

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing stabilizing elements, but do not terminate transcription. Cell **51:** 1145–1157

Stern DB, Gruissem W (1989) Chloroplast messenger RNA 3′ end maturation is biochemically distinct from prokaryotic messenger RNA processing. Plant Mol Biol **13:** 615–626

Stern DB, Jones H, Gruissem W (1989) Function of plastid messenger RNA 3′ inverted repeats: RNA stabilization and gene-specific protein binding. J Biol Chem **264:** 18742–18750

Stern DB, Radwanski ER, Kindle K (1991) A 3′ stem/loop structure of the *Chlamydomonas* chloroplast *atpB* gene regulates mRNA accumulation *in vivo*. Plant Cell **3:** 285–297

Taylor WC (1989) Regulatory interactions between nuclear and plastid genomes. Annu Rev Plant Physiol Plant Mol Biol **40:** 211–233

Tobin EM, Silverthorne J (1985) Light regulation of gene expression in higher plants. Annu Rev Plant Physiol **36:** 569–593

Ye L, Li Y, Fukami-Kobayshi K, Go M, Konishi T, Watanabe M, Sugiura M (1991) Diversity of ribonucleoprotein family in tobacco chloroplasts: two new chloroplast ribonucleoproteins and a phylogenetic tree of ten chloroplast RNA-binding domains. Nucleic Acids Res **19:** 6485–6490