Plant Molecular Biology 4: 39–54 (1985).
© Martinus Nijhoff/Dr W. Junk Publishers, Dordrecht. Printed in the Netherlands.

Multiple transcripts for higher plant *rbc*L and *atp*B genes and localization of the transcription initiation site of the *rbc*L gene

John E. Mullet[1], Emil M. Orozco, Jr. & Nam-Hai Chua
Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, U.S.A.
[1] *Present address: Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843, U.S.A.*

Keywords: *atp*B, chloroplast gene structure, mRNA heterogeneity, *rbc*L, sequence homologies, transcription initiation

Summary

We have compared the *rbc*L and *atp*B transcription units from spinach, maize, and pea. In most cases multiple transcripts were found for a given chloroplast gene. The 5′ termini of these transcripts were determined by S1 nuclease protection and primer extension analyses. The *rbc*L transcripts have 5′ termini 178–179 and 64 nucleotides (spinach), 300 and 59–63 nucleotides (maize), and 178 and 65 nucleotides (pea) upstream from their respective protein coding regions. The *atp*B transcripts have 5′ termini (453–454, 272–273, 179, and 99 nucleotides (spinach), 298–302 nucleotides (maize), and 351–355 nucleotides (pea) upstream from their respective protein coding regions. The intergenic distance between the *rbc*L and *atp*B genes is relatively constant (152 to 157 base pairs) among the three chloroplast genomes. In spinach, maize, and pea, the 80 base pairs surrounding the 5′ end of the *rbc*L gene (±40 base pairs) have 85% sequence homology. Similarly, the 60 base pairs preceding the *atp*B gene have 48% sequence homology. Both genes have '–10' and '–35' regions that resemble the prokaryotic consensus promoter sequence. The larger, but not smaller, *rbc*L transcripts from spinach and pea can be labeled with alpha-^{32}P-GTP by guanylyltransferase. These data suggest that DNA sequences 178–179 (spinach), 300 (maize), and 178 (pea) base pairs before the *rbc*L protein coding regions represent sites of transcription initiation. The sequences 59–65 base pairs before the *rbc*L protein coding regions may correspond to sites of RNA cleavage.

Introduction

Ribulose - 1,5 - bisphosphate carboxylase (RUBISCO) catalyzes the fixation of CO_2 in chloroplasts and is often the most abundant protein in leaf cells. The RUBISCO holoenzyme in most photosynthetic organisms is composed of eight copies of a chloroplast-encoded large subunit of 50–60 kilodaltons and eight copies of a nuclear-encoded small subunit of 12–20 kilodaltons (16, 19). The genes for the large and small subunits of RUBISCO are referred to as *rbc*L and *rbc*S respectively (13). The biosynthesis of *rbc*L and *rbc*S polypeptides is coordinated (16, 40, 42) except under certain conditions (1, 8, 9). In addition, the absolute level of RUBISCO is regulated in response to light (36, 45), cytokinins (9), and in some algae by carbon availability (3). A special example of the regulation of RUBISCO expression occurs in leaf cells of C_4 plants. Mesophyll cells of C_4 plants, which trap CO_2 in the form of C_4 acids, are reported to lack RUBISCO (15, 19) and *rbc*L mRNA (24), whereas bundle sheath cells of C_4 plants fix CO_2 via the Calvin cycle and contain RUBISCO (15, 19) and *rbc*L mRNA (24). Although the *rbc*L gene has been mapped on the chloroplast genome of several plants and sequenced (48), the mechanism by which its expression is regulated in C3 and C4 chloro-

plasts has not been elucidated.

With a few exceptions, DNA sequences that regulate gene expression are found immediately 5′ to the transcription start site (34, 37, 44). Therefore, accurate determination of transcription initiation sites will provide the first step for identifying gene regulatory regions. S1 nuclease protection experiments revealed that the 5′ terminus of the maize *rbc*L mRNA occurs approximately 65 nucleotides upstream of the protein coding region (28). The sequence immediately 5′ to the '–65' position* does not contain any prokaryotic-like '–10' and '–35' promoter elements (28). In contrast, the 5′ termini of the spinach (49) and tobacco (39) *rbc*L transcripts have been reported to be at position '–180'. The chloroplast DNA sequences immediately 5′ to the '–180' position of the spinach and tobacco *rbc*L genes share excellent homology with each other and with the '–10' and '–35' regions of the prokaryotic consensus promoter sequence (39).

In maize (20), spinach (50), and tobacco (38, 41) chloroplast DNAs, the *rbc*L gene is within one kilobase pair of the gene coding for the beta subunit of the chloroplast ATPase. In these plants the genes for the beta and epsilon subunits of the ATPase (*atp*B, *atp*E) are cotranscribed, and the direction of these genes is opposite to that of the *rbc*L gene. Northern hybridization analysis of spinach chloroplast RNA has revealed the presence of at least three *atp*B transcripts (50), but their 5′ termini have not been determined. On the other hand, maize (20) and tobacco (38) chloroplasts have been reported to contain only a single *atp*B transcript. The sequences preceding the 5′-termini of the transcripts contain some homology with the *E. coli* consensus promoter sequence, but not with each other.

The apparent differences in the transcription of the maize, spinach, and tobacco *rbc*L and *atp*B genes raise two intriguing possibilities: (a) the plastid transcriptional apparatus is different between the monocot (maize) and the dicots (spinach and tobacco), and (b) sequence differences surrounding the transcription initiation sites may reflect different regulatory features required by C4 and C3 plant chloroplasts. To investigate these possibilities we reexamined the location of the 5′ termini of the spinach, pea, and maize *rbc*L and *atp*B mRNAs by S1 nuclease protection and primer extension analyses. In this report we confirm previously mapped 5′ termini and identify the 5′ ends of additional transcripts. In particular, we found a larger maize *rbc*L transcript that has sequence homology with the previously identified spinach '–180' *rbc*L transcript. In addition, a smaller transcript of the spinach *rbc*L genes exists that is analogous to the previously identified maize *rbc*L transcript. Two comparable RNAs are also transcribed from the pea *rbc*L gene *in vivo*. Only the larger of the two *rbc*L transcripts from spinach and pea can be radioactively labeled with alpha-^{32}P-GTP by use of guanylyltransferase. Based on these results and DNA homologies at the 5′ ends of the spinach, pea, and maize *rbc*L genes, we conclude that the larger *rbc*L transcript defines a site of transcription initiation, and the smaller *rbc*L transcript is derived from site-specific cleavage of a larger RNA.

* '–65' refers to 65 nucleotides 5′ to the ATG translation initiation codon.

Materials and methods

Reagents

Ribonucleotides and deoxyribonucleotides were from P-L Biochemicals, Inc. Gamma-^{32}P-ATP and alpha-^{32}P-GTP were from New England Nuclear or Amersham Corp. The nucleic acids used as carriers were *E. coli* tRNA type XXI and salmon sperm DNA Type III from Sigma Chemical Co. Trace proteins were removed from the nucleic acids by phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol) extractions, and the nucleic acids were stored in H_2O at –20 °C. In addition, the salmon sperm DNA was sonicated and denatured. ϕX174-*Hae*III DNA fragments were from New England Biolabs Inc. Analytical polyacrylamide gels were prepared with acrylamide (Chemical Dynamics Corp.) and practical grade N,N′-methylenebisacrylamide (Eastman Kodak Co.) that had been treated with activated charcoal (Sigma Chemical Co.) for one h and twice filtered before use. Preparative acrylamide gels were prepared with electrophoresis purity acrylamide (Bio-Rad Laboratories) and N,N′-methylenebisacrylamide (Bethesda Research Laboratories, Inc.). Electrophoresis grade agarose, low melting point agarose, enzyme-grade urea and nucleic acid-grade formamide were also from BRL, Inc. Formamide was deionized for 15 min by treatment with AG 501-X8(D) analytical grade mixed

bed resin and was stored at −20 °C. NA-45 membranes for nucleic acid isolation were from Schleicher and Schuell Co.

Enzymes

Restriction endonucleases, T4 polynucleotide kinase and T4 DNA ligase were from BRL, Inc., New England Biolabs, Inc., and Boehringer-Mannheim. Calf intestine alkaline phosphatase was from Boehringer-Mannheim. S1 nuclease, AMV reverse transcriptase and guanylyltransferase were from BRL, Inc.

Plant growth conditions

Spinacia oleracea (hybrid No. 424), *Zea mays* (FR9 × FR37) and *Pisum sativum* (Progress No. 9) were purchased from the Ferry Morse Seed Co., The Illinois Seed Foundation, and Burpee Seed Co., respectively. Plants were grown in vermiculite in a Conviron environmental growth chamber under the following growth conditions: for spinach, (day 12 h; 22 °C; night 12 h, 18 °C); maize, (day 16 h, 28 °C; night 8 h, 22 °C); and pea, (day 16 h, 22 °C; night 8 h, 18 °C). Plants were harvested 7–10 days after imbibition of dry seeds.

Chloroplast and nucleic acid isolation

Intact chloroplasts were isolated after centrifugation through percoll gradients (2). Plastid DNAs were isolated from spinach, maize, and pea chloroplasts by centrifugation through CsCl-ethidium bromide gradients, similar to the method of Hallick *et al.* (12). To obtain chloroplast, RNA purified plastids were lysed with an equal volume of 6 M urea, 0.36 M NaCl, 1% sodium dodecylsulfate, 20 mM EDTA, 10 mM Tris-HCl, pH 8, and then extracted 3–4 times with phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol). Maize whole cell RNA from purified bundle sheath cells (4) was obtained from Dr. R. Broglie.

Isolation of DNA restriction fragments

DNAs were digested with the appropriate restriction enzymes and then separated by electrophoresis through either agarose, low melting point agarose, or polyacrylamide gels. Restriction fragments were eluted from the gel either by electroelution into dialysis bags (26), electroelution onto NA-45 membranes (25), the 'crush and soak' procedure of Maxam & Gilbert (27), or by solubilizing the low melting point agarose (5 min at 70 °C) and purifying the DNA with heated (55 °C) DE-52 column chromatography (31). DNAs isolated by electroelution were further purified by DE-52 column chromatography at room temperature.

Plasmid DNA constructions

The DNA restriction fragments containing the 5′ ends of the *rbc*L and *atp*B genes from spinach (49), pea (32), and maize (28) plastid DNAs have been identified previously. The appropriate fragments were isolated from agarose gels. Plasmid DNAs pSoc801 and pSoc802 contain the 1041 bp *Msp*I-*Pst*I fragment of spinach plastid DNA inserted into the *Acc*I and *Pst*I sites of pUC8 and pUC9 (46), respectively. Plasmid DNAs pPsc650 and pPsc660 contain the 1710 bp *Bam*HI-*Pst*I fragment of pea plastid DNA inserted into the *Bam*HI and *Pst*I sites of pUC8 and pUC9, respectively. The plasmid DNA pZmc800 contains the 953 bp *Xma*I-*Pst*I fragment of maize plastid DNA inserted into the *Xma*I and *Pst*I sites of pUC8. Plasmid DNAs, transformed into bacterial strains JM103 (pSoc801, pSoc802) or JM83 (pPsc650, pPsc660, pZmc800), were isolated after centrifugation on cesium-ethidium bromide gradients (6).

Preparation and sequence determination of radioactively labeled DNAs

Linear DNA was 5′ end labeled with gamma ^{32}P-ATP and T4 polynucleotide kinase (27). The ^{32}P-DNA was denatured and the complementary strands separated on 5% polyacrylamide gels. For some experiments, the ^{32}P-DNA was restricted and the two double-stranded ^{32}P-DNAs were separated on 5, 8, or 12% polyacrylamide gels. Initially, the ^{32}P-DNAs were purified from acrylamide gels by either electroelution into dialysis bags (26) or by the 'crush and soak' procedure of Maxam & Gilbert (27). Of these two methods, the 'crush and soak' eluted ^{32}P-DNAs were found to be more reliable for use in subsequent experiments. Electroeluted ^{32}P-DNAs frequently became insoluble after precipitation from ethanol, perhaps due to a contaminant from the dialysis tubing. DNA sequencing reactions were performed according to the method of Maxam & Gilbert (27).

S1 nuclease protection assay

S1 nuclease protection assays were performed essentially as described by Weaver & Weissman (47). The 5′ end labeled ^{32}P-DNA (2 500 to 20 000 cpm/

assay) and carrier *E. coli* tRNA (10 μg/assay) were resuspended in 80% formamide (10 μl/assay), boiled for 5 min, and then added to a dry pellet containing chloroplast RNA (0.03 to 15 μg/assay) and S1 hybridization buffer to give a final concentration of 400 mM NaCl, 40 mM Pipes-NaOH, pH 6.4, 1 mM EDTA. Hybridization was performed for 6–18 h. For single-stranded DNAs the temperature of hybridization was 37 °C, whereas for double-stranded DNAs the temperature was often optimized (37°–57 °C) to minimize DNA:DNA renaturation (47). After hybridization, 25 units of S1 nuclease diluted into 0.1 ml of buffer was added (250 mM NaCl, 30 mM NaOAc, pH 4.6, 1 mM $ZnSO_4$, 20 μg/ml denatured salmon sperm DNA), and the reaction was incubated at 30 °C for 40 min. The S1-protected nucleic acid was precipitated twice with ethanol, using 5 μg of *E. coli* tRNA as carrier, and the ^{32}P-DNAs were analyzed on 6% polyacrylamide-8.3 M urea DNA sequencing gels (27).

Primer extension assay

A 5′ end labeled double-stranded ^{32}P-DNA fragment was denatured and hybridized to chloroplast RNA as described for the S1 nuclease protection assay. After hybridization the nucleic acid was precipitated from ethanol once and resuspended in 20 μl of 50 mM Tris-HCl, pH 8.3, 70 mM KCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 0.1 mg/ml BSA, and 1 mM of each dNTP. Primer extension was initiated with the addition of 20–50 units of AMV reverse transcriptase, incubated at 37 °C for 60 min, and then stopped with the addition of 0.13 ml of 6 M urea, 0.36 M NaCl, 1% sodium dodecylsulfate, 20 mM EDTA, 10 mM Tris-HCl, pH 8. The reaction mixture was extracted twice with 0.15 ml of phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol). The nucleic acids were precipitated twice from ethanol and then analyzed on 6% polyacrylamide-8.3 M urea DNA sequencing gels (27).

Analysis of RNA by use of guanylyltransferase

Chloroplast RNAs from pea and spinach were 5′ end labeled with alpha-^{32}P-GTP and guanylyltransferase essentially as described by Levens *et al.* (23). A 20 μl reaction contained the following components: 3–20 μg of chloroplast RNA, 10 units of vaccinia virus guanylyltransferase, 0.10–0.15 mCi of alpha-^{32}P-GTP, 50 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$, and 1 mM dithiothreitol. After incubation at 37 °C for 30 min the reaction was stopped and extracted with phenol:chloroform:isoamyl alcohol as described above for primer extended DNAs. The resulting chloroplast ^{32}P-RNA (2 500–25 000 cpm/assay) was hybridized to unlabeled double- or single-stranded DNAs (15 to 150 fmoles/assay) and treated with S1 nuclease as described above. Before analysis on denaturing polyacrylamide gels the S1 nuclease protected ^{32}P-RNA:DNA hybrids were denatured by heating at 80 °C (instead of 100 °C) for 5 min. In our hands, the extent of *in vitro* capping was not increased by prior treatment of the chloroplast RNA with glyoxal.

Results

DNA sequence of the spinach rbc*L*-atp*B intergenic region*

Restriction maps of the cloned plastid DNAs from spinach, pea, and maize are shown in Fig. 1. Restriction sites used for the preparation of DNA fragments for S1 nuclease protection, primer extension, and DNA sequencing experiments are also indicated. The DNA sequence of *rbc*L-*atp*B intergenic region from spinach was determined by the Maxam-Gilbert technique (27) and is presented in Figs. 7 and 10. Much of this sequence has been confirmed independently (C. Poulsen & N.-H. Chua, unpublished observations) by the dideoxy sequencing method (35). The DNA sequence of the spinach *rbc*L-*atp*B intergenic region has been published recently (7). The sequence presented in Fig. 10A differs from the previously reported sequence at position '–3' where we have found the nucleotide adenosine instead of guanosine. We note that the presence of guanosine at this position would generate a *Bst*NI restriction site (5′-CCAGG) which is not present in our pSoc801 or pSoc802 plasmid DNAs. The sequence of the corresponding region from maize (20) and pea (G. Zurawski, personal communication) have also been determined.

The spinach rbc*L gene has a '–65' transcript*

The 5′ end of the maize *rbc*L transcript has been mapped by McIntosh *et al.* (28) to position

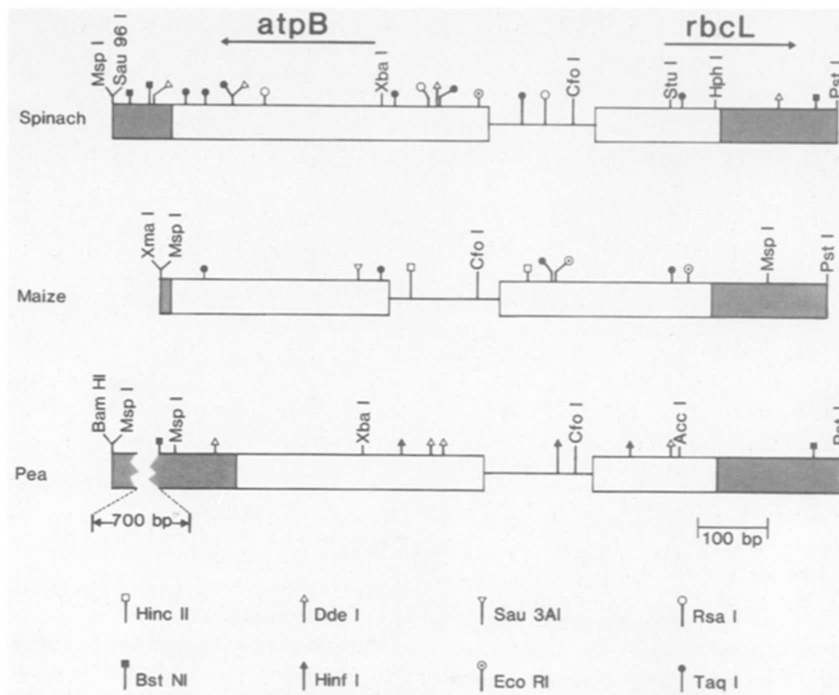

Fig. 1. Restriction enzyme map of *rbc*L-*atp*B region from spinach, maize and pea. Restriction maps are shown for *rbc*L-*atp*B region from spinach (1041 bp *Msp*I-*Pst*I DNA), maize (953 bp *Xma*I-*Pst*I DNA) and pea (1710 bp *Bam*HI-*Pst*I DNA). These chloroplast DNA fragments were purified and plasmid DNAs constructed as described in the text. Various restriction fragments were isolated from these plasmid DNAs for use in S1 nuclease protection, primer extension and DNA sequence analyses. The restriction maps shown above are for the enzymes used in this report. The protein coding region of each gene is represented as a filled-in box. The transcribed but untranslated regions are shown as open boxes. The boundary of each open box represents the size of the largest transcript of that gene (see text). To obtain the DNA sequence of the spinach intergenic region we performed Maxam-Gilbert DNA sequencing reactions (27) on the 510 bp *Eco*RI-*Pst*I DNA (5′ end labeled at the *Eco*RI site) and the two strands of the 226 bp *Taq*I fragment. The DNA sequence obtained is shown in Figs. 7 and 10A.

59–63 bp before the protein coding region. While the sequences of the spinach and maize chloroplast DNAs in this region (positions '–1' to '–136') are 69% homologous, no spinach *rbc*L transcript was identified at this locus. Instead, the 5′ end of the spinach *rbc*L transcript was mapped by Zurawski *et al.* (49) to position '–180'. The DNA probe used by Zurawski *et al.* (49) for their S1 nuclease protection experiment was a 226 nt single-stranded *Taq*I fragment labeled at the '–58' position. Their S1 data were confirmed with reverse transcription analysis, using as primer a 14 nt *Taq*I-*Hae*III DNA labeled at the *Taq*I site (position '–58'). Because of the choice of the S1 probe and the reverse transcription primer, a shorter '–65' transcript, if present, might have escaped detection. To check this possibility we repeated the S1 nuclease protection analysis using a 952 nt single-stranded *Bst*NI fragment that spans the entire region separating the *rbc*L and *atp*B protein coding loci (Fig. 1). Conditions were varied to optimize the signal. In particular, the RNA:DNA ratio was varied over three orders of magnitude, and part of this titration is shown in Fig. 2. At high RNA:DNA ratios the primary RNA detected has its 5′ terminus at position '–180' (Fig. 2, lane 1), confirming the results of Zurawski *et al.* (49). At lower RNA:DNA ratios, however, we found a second *rbc*L transcript with 5′ terminus at position '–65' (Fig. 2, lane 2). A potential artifact on the S1 nuclease protection assay is cleavage of RNA:DNA hybrids at AT-rich sequences (14). To rule out such an artifact, we confirmed the 5′ end assignments by primer extension analysis (Fig. 2, lane 3). The primer extension assay can also suffer from artifacts

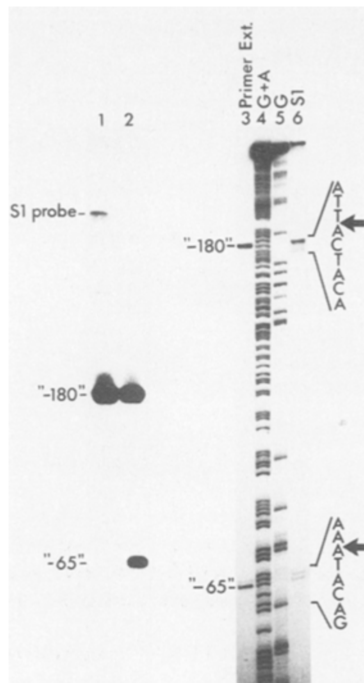

Fig. 2. The 5′ end analysis of spinach *rbc*L mRNAs. A 952 bp *Bst*NI DNA fragment was 5′ end labeled with gamma-^{32}P-ATp and T4 polynucleotide kinase and the *rbc*L coding strand was isolated. This *Bst*NI ^{32}P-DNA (10 000 cpm/reaction) was hybridized with either 7.2 μg (lane 1) or 0.72 μg (lane 2) of spinach chloroplast RNA and then incubated with S1 nuclease. The mobilities of the 952 nt *Bst*NI DNA and the DNAs protected by '–180' and '–65' *rbc*L mRNAs are indicated to the left of lane 1.
Lane 3: 15 000 cpm of the 88 bp *Dde*I-*Hph*I DNA (5′ end labeled at the *Dde*I site) was denatured, hybridized with 0.5 μg of spinach chloroplast RNA and extended with AMV reverse transcriptase.
Lanes 4 and 5: Maxam-gilbert 'G + A' and 'G' DNA sequencing reactions, respectively, of the 334 bp double-stranded *Dde*I-*Rsa*I DNA (5′ end labeled at the *Dde*I site).
Lane 6: 15 000 cpm of the 334 b *Dde*I-*Rsa*I DNA was denatured, hybridized with 30 ng of spinach chloroplast RNA and then digested with S1 nuclease. The mobilities of the primer extended DNAs relative to the DNA sequence are shown to the right of lane 6.

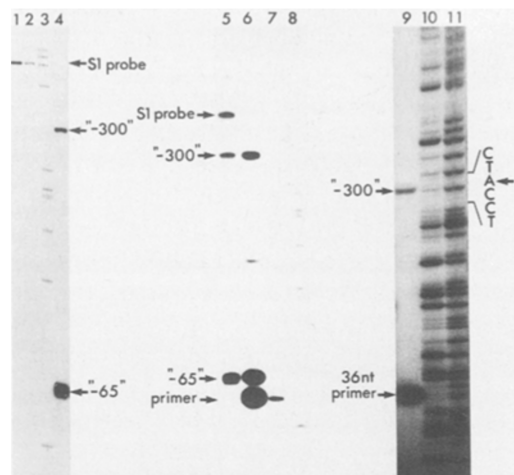

Fig. 3. The 5′ end analysis of the maize *rbc*L mRNAs.
Lanes 1 and 2: the purified *rbc*L coding strand of the 869 bp *Msp*I DNA fragment.
Lanes 3 and 8: the 5′ end labeled ϕX174-*Hae*III DNA fragments used as size standards.
Lane 4: an excess of the purified *Msp*I single-stranded DNA was hybridized with total RNA from maize bundle sheath cells and then incubated with S1 nuclease. The DNAs protected from S1 nuclease by '–300' and '–65' *rbc*L mRNAs are indicated to the right of lane 4.
Lane 5: the 583 bp *Msp*I-*Sau*3AI DNA (5′ end labeled at the *Msp*I site) was denatured, hybridized with maize bundle sheath RNA and then digested with S1 nuclease.
Lane 6: the 135 bp *Msp*I-*Taq*I DNA (5′ end labeled at the *Msp*I site) was denatured, hybridized with the maize bundle sheath RNA and extended with AMV reverse transcriptase.
Lane 7: the 135 bp *Msp*I-*Taq*I DNA used for the adjacent primer extension experiment.
Lane 9: the 36 bp *Eco*RI:*Hinc*II DNA (5′ end labeled at the *Eco*RI site) was denatured, hybridized with maize bundle sheath RNA and extended with AMV reverse transcriptase.
Lanes 10 and 11: Maxam-Gilbert's 'G' and 'G + A' DNA sequencing reactions, respectively, of the 566 bp *Eco*RI-*Xma*I DNA (5′ end labeled at the *Eco*RI site). The mobility of the primer extended DNA with respect to the maize plastid DNA sequence is shown to the right of lane 11.

such as premature termination (22). Bands that were not detected by both S1 nuclease protection and primer extension analyses were assumed to be artifactual. The position of each 5′ terminus with respect to the plastid DNA sequence is also shown in Fig. 2 (lanes 3–6).

The maize rbcL *gene has a '–300' transcript*

The detection of a smaller '–65' *rbc*L transcript in spinach raised the possibility that a larger maize *rbc*L transcript might also have been overlooked. In their S1 nuclease protection experiment McIntosh *et al.* (28) used as a probe a 443 bp *Hinf*I-*Hinc*II fragment labeled at the *Hinf*I site (position '+181'). This DNA probe would have detected mRNA termini occurring up to 262 nt 5′ to the *rbc*L

protein coding region. For our experiments we used a 869 bp *Msp*I DNA fragment from pZmc800 that spans the entire region separating the maize *rbc*L and *atp*B protein coding loci (Fig. 1). The isolated *rbc*L coding strand of this DNA was hybridized to whole cell RNA extracted from bundle sheath preparations. After S1 nuclease digestion two *rbc*L transcripts were identified with 5′ ends at positions '–300' and '–65' (Fig. 3, lane 4). No full-length protection of the single-stranded DNA probe was detected, demonstrating the absence of any larger transcripts for the maize *rbc*L gene. Although the *rbc*L '–65' transcript had been identified previously by S1 nuclease protection experiments (28), the larger *rbc*L transcript had not been noted. For the S1 nuclease protection experiment shown in lane 5, we used a double-stranded 603 bp *Msp*I-*Sau*3AI DNA fragment. The 5′ ends at positions '–65' and '–300' are indicated, and some full-length protection of the DNA probe is observed, presumably due to DNA:DNA renaturation. In addition, the positions of the two *rbc*L 5′ ends were confirmed by primer extension analysis (Fig. 3, lane 6). The approximately 220 nt ^{32}P-DNA shown in lane 6 is presumed to be a primer extension artifact. We have also detected the presence of two maize *rbc*L transcripts by Northern hybridization analysis (data not shown). McIntosh *et al.* (28) mapped the 5′ end of the '–65' *rbc*L transcript by S1 protection analysis to position '–59' to '–63'. We have mapped the 5′ end of the '–300' *rbc*L transcript by primer extension analysis to position 300 nt before the *rbc*L protein coding region (Fig. 3, lanes 9–11).

The pea rbc*L gene has both '-180' and '–65' transcripts*

To determine if the presence of the two *rbc*L transcripts is a common occurrence in higher plant chloroplasts, we extended our analysis to the pea *rbc*L gene. The 932 bp *Bst*NI DNA fragment from pPsc650 spans the entire region separating the *rbc*L and *atp*B protein coding loci (Fig. 1). The two strands of this DNA fragment did not readily separate after polyacrylamide gel electrophoresis. Therefore, the DNA probe used for S1 nuclease protection analysis was a double-stranded DNA restriction fragment in which only one 5′ end (at the *Bst*NI site) was radioactively labeled. The use of a double-stranded DNA fragment for S1 nuclease protection resulted in some full-length protection

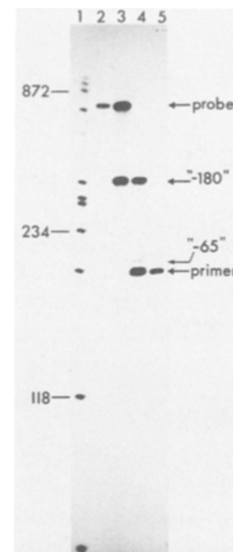

Fig. 4. The 5′ end analysis of the pea *rbc*L mRNAs.
Lane 1: 5′ end labeled ϕX174-*Hae*III DNA fragments. The sizes (bp) of three of these size standards are shown to the left of lane 1.
Lane 2: the 644 bp *Xba*I-*Bst*NI DNA fragment (5′ end labeled at the *Bst*NI site).
Lane 3: the *Xba*I-*Bst*NI DNA and pea chloroplast RNA were hybridized and then incubated with S1 nuclease.
Lane 4: the 189 bp *Acc*I-*Bst*NI DNA fragment (5′ end labeled at the *Bst*NI site) was hybridized with pea chloroplast RNA and then extended by AMV reverse transcriptase.
Lane 5: The 189 bp *Acc*I-*Bst*NI DNA.

of the DNA probe due to renaturation of the two DNA strands (Fig. 4, lane 3). Alternatively, this signal could have been due to full protection of the coding strand by an RNA that is longer than the probe. If this were the case, such an RNA would have resulted in a ^{32}P-DNA larger than the S1 probe after primer extension analysis, but this was not observed (Fig. 4, lane 4). The major pea *rbc*L transcript has its 5′ end approximately 180 nt before the *rbc*L protein coding region. Upon larger exposure of the autoradiogram, a minor transcript with 5′ terminus at the '–65' position is also detected by both S1 and primer extension analyses. The DNA sequence of the pea *rbc*L gene and the precise location of the '–180' 5′ end have been determined recently by G. Zurawski (personal communication).

Identification of the transcription initiation sites for the spinach and pea rbc*L genes*

The presence of two *rbc*L mRNAs raises the question of whether these species are derived from

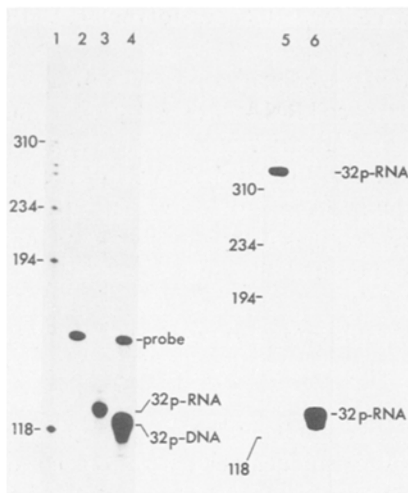

Fig. 5. The transcription initiation site for the pea *rbc*L gene is at position '-180'. DNA fragments were 5′ end labeled with gamma-^{32}P-ATP, and T4 polynucleotide kinase and pea chloroplast RNA was 5′ end labeled with alpha-^{32}P-GTP and guanylyltransferase.
Lane 1: 5′ end labeled ϕX174;*Hae*III DNA fragments. Molecular size standards (bp) are indicated to the left of lanes 1 and 5.
Lane 2: the 151 bp *Acc*I-*Cfo*I DNA fragment (5′ end labeled at the *Acc*I site).
Lanes 3 and 6: the 5′ end labeled pea chloroplast RNA was protected from S1 nuclease digestion by prior hybridization to unlabeled 151 bp *Acc*I-*Cfo*I DNA.
Lane 4: the 5′ end labeled *Acc*I-*Cfo*I ^{32}P-DNA was hybridized with unlabeled pea chloroplast RNA prior to S1 treatment.
Lane 5: the 5′ end labeled pea chloroplast RNA was hybridized with a 644 bp unlabeled *Bst*NI-*Xba*I DNA before S1 treatment.

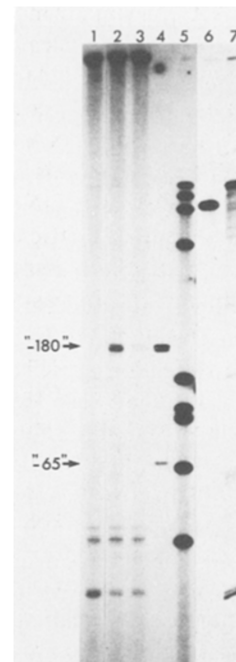

Fig. 6. Transcription of the spinach *rbc*L gene is initiated *in vivo* at position '-180'. DNA was 5′ end labeled with gamma-^{32}P-ATP and T4 polynucleotide kinase. Spinach chloroplast RNA was 5′ end labeled with alpha-^{32}P-GTP and guanylyltransferase.
Lane 1: spinach chloroplast ^{32}P-RNA was digested with S1 nuclease without prior hybridization to DNA.
Lane 2: the purified unlabeled 1052 nt *Bam*HI-*Hind*III DNA from pSoc801 (the *rbc*L coding strand) was hybridized with spinach chloroplast ^{32}P-RNA and then digested with S1 nuclease.
Lane 3: the 1052 bp double-stranded *Bam*HI-*Hind*III DNA from pSoc801 was denatured and hybridized to spinach chloroplast ^{32}P-RNA prior to S1 nuclease digestion.
Lane 4: the purified 1052 nt single-stranded *Bam*HI-*Hind*III ^{32}P-DNA (the *rbc*L coding strand) was hybridized to unlabeled spinach chloroplast RNA and then incubated with S1 nuclease.
Lane 5: 5′ end labeled ϕX174-*Hae*III DNA fragments.
Lane 6: the 1052 nt single-stranded *Bam*HI-*Hind*III ^{32}P-DNA (the *rbc*L coding strand) without S1 treatment.
Lane 7: spinach chloroplast ^{32}P-RNA without S1 treatment.
Arrows indicate positions where radioactively labeled nucleic acids should migrate if S1 nuclease protection is due to a '-180' or '-65' *rbc*L mRNA.

independent transcription initiation events. To investigate this possibility we have made use of the enzyme guanylyltransferase, which catalyzes the addition of GMP to the 5′ end of an RNA containing a 5′ terminal polyphosphate (29). The initial transcription product of a gene has a 5′ terminal triphosphate, whereas an RNA which is derived from processing of a primary transcript would have either a monophosphate or a hydroxyl group at its 5′ end (11). The ability of guanylyltransferase to add GMP to the 5′ end of an RNA, therefore, defines a site of transcription initiation of that gene.

Pea chloroplast RNA was radioactively labeled by incubation with alpha-^{32}P-GTP and guanylyltransferase. The product was hybridized to a 151 bp *Acc*I-*Cfo*I or to a 644 bp *Bst*NI-*Xba*I DNA fragment, and the resulting hybrids were digested with S1 nuclease. The protected ^{32}P-RNAs are approximately the same size as the complementary ^{32}P-DNAs that are produced by standard S1 nuclease protection assays, demonstrating that the '-180' species is a primary transcript (Fig. 5, lanes 3, 5, and 6). The ^{32}P-RNA migrates slightly slower than the complementary ^{32}P-DNA (Fig. 5, lanes 3 and 4). No labeled RNA was detected corresponding to the '-65' *rbc*L transcript. This transcript is a minor species in pea. Therefore, we analyzed the spinach

chloroplast RNA in a similar manner (Fig. 6). As a reference, a standard S1 nuclease protection assay was performed using unlabeled spinach chloroplast RNA and the isolated *rbc*L coding strand of the 1052 bp *Bam*HI-*Hind*III DNA from pSoc801, labeled at the *Hind*III site. This ^{32}P-DNA contains seven nucleotides of vector DNA sequence, but is still protected, although inefficiently, from S1 nuclease digestion (lane 4). To assay the *rbc*L primary transcripts, spinach chloroplast ^{32}P-RNA was hybridized to the double-stranded 1052 bp *Bam*HI-*Hind*III DNA fragment. A faint signal is seen due to S1 nuclease protection of the '–180' transcript (lane 3). This signal is amplified if the ^{32}P-RNA is hybridized, instead, to the purified *rbc*L coding strand of the *Bam*HI-*Hind*III fragment (lane 2). In the pea experiment, the ^{32}P-RNA migrated slightly slower than the complementary ^{32}P-DNA. In the spinach experiment, the protected ^{32}P-DNA is seven nucleotides larger than the complementary ^{32}P-RNA, resulting in the comigration of the two '–180' signals (lanes 2 and 4). No signal was observed that would have corresponded to a ^{32}P-GMP labeled '–65' transcript (lane 2). These results indicate that the '–180', but not the '–65', mRNA is a primary transcript.

When maize chloroplast RNA and maize bundle sheath total cell RNA were analyzed by this method, all the radioactivity was retained as an aggregate at the top of the gel. Neither of the two maize *rbc*L transcripts were detected. This technical difficulty prevented us from identifying the maize *rbc*L primary transcript.

Sequence homology at the 5′ end of higher plant rbc*L genes*

The above results indicate that transcription initiation of the spinach and pea *rbc*L genes occurs at the '–180' and not the '–65' position. The DNA sequences from the spinach and maize *rbc*L '–65' regions have been compared (49). These two regions from position '–1' to '–136' contain 69% sequence homology and are not homologous with the *E. coli* consensus promoter sequence (49). The DNA sequences surrounding the *rbc*L '–180' regions from spinach and tobacco are highly homologous to each other and to the *E. coli* consensus

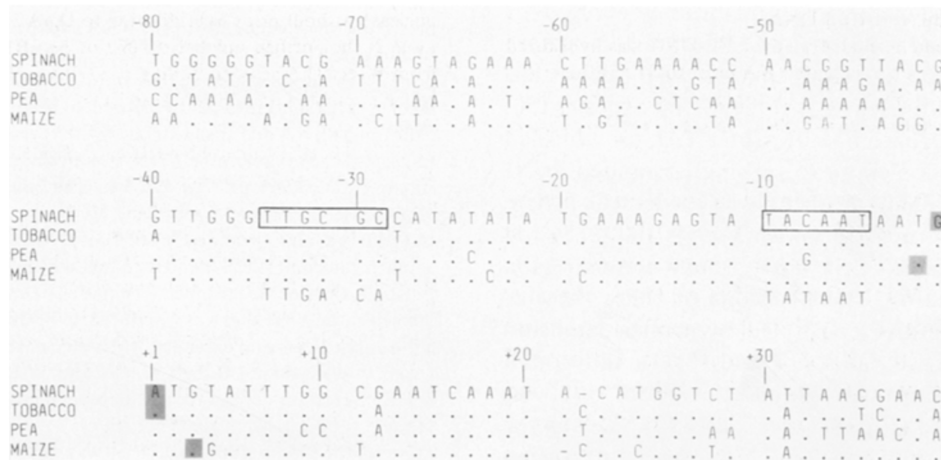

Fig. 7. DNA sequence homology at the 5′ ends of *rbc*L genes. The DNA sequences at the 5′ end of the *rbc*L genes from spinach (49; this report), tobacco (39), pea (G. Zurawski, personal communication) and maize (20) are shown. The maize plastid DNA fragment spanning positions '–65' (*Aha*III site) to '+35' (*Hpa*I site) has recently been cloned and the sequence determined (Hanley-Bowdoin L, Orozco EM, Jr, Chua, N-H, unpublished observations). This newly determined sequence differs in one nucleotide from the published sequence (20) at position '+20' (5′-TCCACG instead of 5′-TCACG) and is used in this figure. The 5′ end of the '–180' transcripts from spinach (49; this report), tobacco (20), and pea (G. Zurawski, personal communication) and of the '–300' transcript from maize (this report) are indicated by shaded boxes. The *rbc*L 5′ termini identified in this report were determined by subtracting 1.5 bases from the apparent mobilities of the corresponding primer extended DNAs (43). Homology with the spinach DNA sequence at a particular position is indicated by a period whereas absence of a base is indicated by a hyphen. The conserved '–10' and '–35' sequences of the consensus prokaryotic promoter are also shown. The corresponding prokaryotic-like sequences in the spinach plastid DNA are enclosed within boxes.

promoter sequence (39). We have extended this sequence comparison to include the '–180' transcript from pea (G. Zurawski, personal communication) and the '–300' transcript from maize. The sixty base pairs surrounding the 5′ end of the *rbc*L gene from spinach, tobacco, pea, and maize have 82% homology (Fig. 7). Sequences outside these regions retain much less homology. The highly conserved sequence contains prokaryote-like '–10' and '–35' regions. In addition, the sequence conservation ends 40 bp 5′ to the transcription initiation site of the *rbc*L gene.

Determination of the 5′ ends of four transcripts for the spinach atp*B gene*

In maize (20), spinach (50), and tobacco (38) chloroplast DNAs, the *rbc*L and *atp*B coding regions are within one kilobase pair of each other and are transcribed divergently. A relatively short sequence of plastid DNA should, therefore, contain the promoter regions of these two genes. Accurate determination of the transcription start site for the *atp*B gene would provide information on the precise length of this putative regulatory region.

Single *atp*B transcripts were initially reported for maize (20) and tobacco (38). In contrast, Northern hybridization analysis of spinach *atp*B mRNA has revealed the presence of three distinct species (50; Mullet JE, Orozco EM, Jr, Chua N-H, unpublished observations). Using a 952 nt single-stranded *Bst*NI DNA fragment as probe, we have identified three major *atp*B transcripts with 5′ ends 100, 275, and 455 nt upstream of the *atp*B protein coding region (Fig. 8, lane 6). The positions of these three 5′ termini were also confirmed by primer extension analyses (Fig. 8, lanes 1, 7 and 13). In addition, a minor transcript with 5′ end at position '–180' was also observed (Fig. 8, lanes 7 and 12). As was the case with the *rbc*L mRNA, with increasing RNA:DNA ratios the signals due to the smaller *atp*B transcripts decrease. At a sufficiently high RNA:DNA ratio the only apparent S1 nuclease protected DNA is that due to protection by the largest ('–455') *atp*B transcript (data not shown).

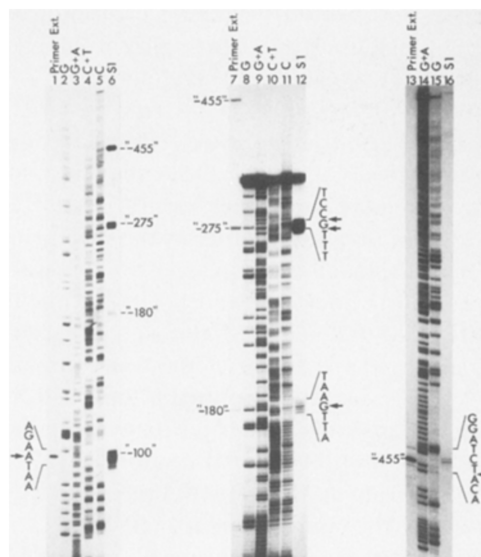

Fig 8. The 5′ end analysis of the spinach *atp*B MRNAs.
Lane 1: the double-stranded 58 bp *Sau*96I-*Sau*3AI DNA (5′ end labeled at the *Sau*96I site) was denatured, hybridized with spinach chloroplast RNA and extended with reverse transcriptase.
Lanes 2–5: Maxam-Gilbert DNA sequencing reactions of the single-stranded 1038 *Sau*96I-*Pst*I DNA (5′ end labeled at the *Sau*96I site).
Lane 6: the single-stranded 1038 nt *Sau*96I-*Pst*I DNA was hybridized with spinach chloroplast RNA and then incubated with S1 nuclease.
Lane 7: the double-stranded 83 bp *Taq*I-*Rsa*I DNA (5′ end labeled at the *Taq*I site) was denatured, hybridized with spinach chloroplast RNA and then extended with reverse transcriptase.
Lanes 8–11: Maxam-Gilbert DNA sequencing analysis of the single-stranded 269 nt *Taq*I DNA (5′ end labeled at the same *Taq*I site as the primer used in lane 7).
Lane 12: the single-stranded 269 nt *Taq*I DNA was hybridized to spinach chloroplast RNA and then incubated with S1 nuclease.
Lane 13: the double-stranded 77 bp *Xba*I-*Dde*I DNA (5′ end labeled at the *Xba*I site) was denatured, hybridized with spinach chloroplast RNA and extended with reverse transcriptase.
Lanes 14–15: Maxam-Gilbert 'G + A' and 'G' DNA sequencing reactions, respectively, of the double-stranded 648 bp *Xba*I-*Pst*I DNA (5′-end labeled at the *Xba*I site).
Lane 16: the 648 bp *Xba*I-*Pst*I DNA was denatured, hybridized with spinach chloroplast RNA, and then incubated with S1 nuclease. Primer extended and S1 nuclease protected DNAs that identify the four 5′ termini of *atp*B mRNAs ('–100', '–180', '–275', '–455') are indicated. The mobilities of the primer extended DNAs relative to the DNA sequence are indicated with arrows.

Determination of the 5′ ends of the maize and pea atp*B transcripts*

A single 5′ end for the maize *atp*B transcript has been mapped by S1 nuclease protection analysis to within position '–298' to '–302' (20). We have confirmed this assignment of the 5′ end by both S1 nuclease protection and primer extension analyses (data not shown). We did not detect any other transcripts for the maize *atp*B gene.

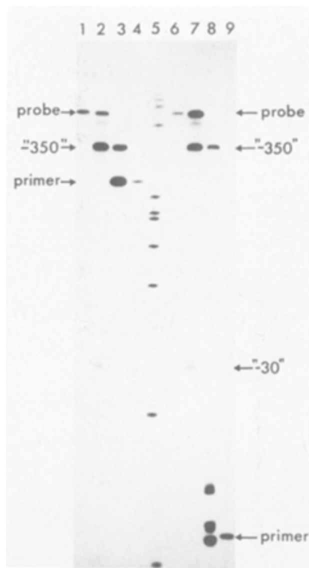

Fig. 9. The 5′ end analysis of the pea *atp*B mRNAs.
Lanes 1 and 6: the 743 bp *Bst*NI-*Acc*I DNA fragment (5′ end labeled at the *Bst*NI site).
Lanes 2 and 7: the *Bst*NI-*Acc*I DNA and pea chloroplast RNA were hybridized and then digested with S1 nuclease.
Lane 3: the 344 bp *Bst*I-*Hin*fI DNA (5′ end labeled at the *Bst*NI site) was hybridized with pea chloroplast RNA and extended by AMV reverse transcriptase.
Lane 4: the 344 bp *Bst*NI-*Hin*fI DNA.
Lane 5: the 5′ end labeled φX174-*Hae*III DNA fragments used as size standards.
Lane 8: the 79 bp *Bst*NI-*Dde*I DNA (5′ end labeled at the *Bst*NI site) was denatured, hybridized with pea chloroplast RNA and extended by AMV reverse transcriptase.
Lane 9: the 79 bp *Bst*NI-*Dde*I DNA. The 5′ end of the pea *atp*B mRNA is indicated ('-350'). An artifactual signal at position '-30' is also detected by S1 nuclease protection analysis (see text).

We have also examined the *atp*B transcripts from pea. For reasons mentioned above, a double-stranded DNA 5′ end labeled at the *Bst*NI site was used for S1 nuclease protection (Fig. 9). As with maize, the pea *atp*B mRNA is homogeneous at the 5′ terminus, which occurs approximately 350 nucleotides before the *atp*B protein coding region (Fig. 9). The DNA sequence of the pea *atp*B gene and the precise location of the 5′ end of the *atp*B transcript have been determined recently by G. Zurawski (personal communication). In Fig. 9, full-length protection of the S1 probe (lanes 2 and 7) is presumably due to DNA:DNA renaturation, since an RNA larger than the probe is not detected by primer extension analysis (lanes 3 and 8). A band corresponding to a 5′ end at the '-30' position, however, is also detected by S1 nuclease protection (lane 7). The intensity of the '-30' signal varies with the particular conditions of the S1 nuclease protection experiment, and is often of greater intensity than the '-350' signal (data not shown). This '-30' signal, however, is not detected by primer extension analysis (lane 8). Fourteen consecutive thymidine residues occur in this region of the pea *atp*B gene (G. Zurawski, personal communication). The '-30' signal that we detect by S1 protection analysis is, therefore, most likely an artifact due to transient 'breathing' of the RNA:DNA hybrid at an A-T rich sequence and subsequent cleavage by S1 nuclease (14).

Sequence homology at the 5′ end of higher plant atp*B genes*

The DNA sequences surrounding the 5′ end of the *atp*B genes from spinach, pea, and maize are presented in Fig. 10A. The largest *atp*B transcript from spinach is used to define the 5′ end of the gene. In addition, the corresponding region from the tobacco plastid DNA is included. These four plastid DNAs retain 48% homology for the sixty base pairs preceding the *atp*B gene. If the monocot maize is exluded from the comparison, the remaining dicot plants retain 78% homology for this sixty bp region, and have a striking 92% sequence homology for the 37 bp 5′ to the presumptive transcription initiation site. The conserved chloroplast DNA sequence shown in Fig. 10A contains the sequence 5′-TTGACA of the '-35' region of the *E. coli* consensus promoter sequence (34). A 5′ end of the tobacco gene has been mapped at another locus (38). The sequence comparison shown in Fig. 10A suggests that an additional larger *atp*B transcript may have initially been overlooked. In a more recent report, two *atp*B transcripts from tobacco have been detected by Northern hybridization experiments (10).

Comparison of the DNA sequences that precede the four 5′ ends of the spinach atp*B mRNAs*

The four 5′ ends identified for transcripts of the spinach *atp*B gene are potentially sites of transcription initiation and/or RNA processing. A comparison of the DNA sequences preceding the four 5′

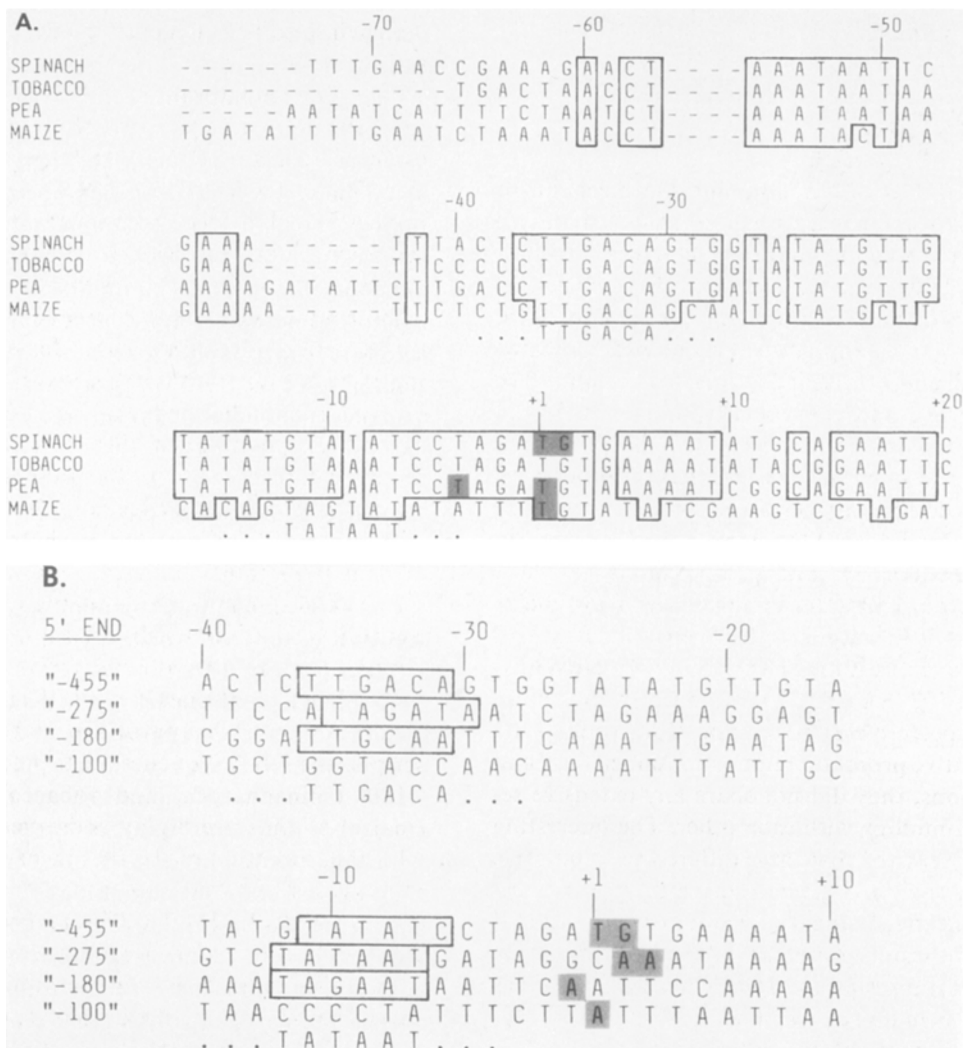

Fig. 10. (A) DNA sequence homology at 5′ ends of *atp*B genes. The DNA sequence surrounding the 5′ end of the putative transcription initiation site ('–455') of the spinach *atp*B gene is compared with the corresponding regions from tobacco (38), pea (G. Zurawski, personal communication), and maize (20). The conserved sequences are boxed. The prokaryotic '–10' and '–35' consensus promoter sequences are also shown. The 5′ ends of three *atp*B genes from spinach '–455'; this report), pea ('–350'; G. Zurawski, personal communication) and maize ('–300'; 20) are indicated by shaded boxes. A 5′ end for the tobacco gene has been mapped at another locus (38). The sequence comparison shown here suggests that an additional larger tobacco *atp*B transcript may have been overlooked. The chloroplast DNA sequences shown here are contiguous with the corresponding chloroplast DNA sequences shown in Fig. 7. (B) DNA sequences surrounding the 5′ ends of the four spinach *atp*B transcripts. The 5′ terminal nucleotides were determined by primer extension experiments (this report) and are indicated with shaded boxes. Each 5′ termini was determined by substrating 1.5 bases from the apparent mobility of primer extended DNA relative to an adjacent sequence ladder (43). The prokaryotic '–10' and '–35' consensus promoter sequences are also shown. Plastid DNA sequences similar to the '–10' and '–35' elements are enclosed in boxes.

termini does not reveal any significant sequence conservation (Fig. 10B). Three of the four 5′ ends have sequence homology with the '–10' and '–35' regions of the *E. coli* consensus promoter sequence. However, there is currently no evidence (e.g. capping of *in vivo* mRNAs) that these regions promote transcription of the *atp*B gene. A 5′ end of the tobacco *atp*B gene has been assigned to position '–257' (38). The sequence shown in Fig. 10B surrounding the spinach '–275' 5′ end is 80% homolo-

gous with the corresponding sequence at the tobacco '–257' position.

Discussion

We are interested in studying the mechanisms that regulate chloroplast gene transcription. In particular, we would like to investigate the differences that may exist in the expression of the *rbc*L and *atp*B genes from C3 (e.g. spinach, pea, and tobacco) and C4 (e.g. maize) plants. Early reports indicated that the 5′ end of the *rbc*L genes from maize (28), spinach (49), and tobacco (39) occurred 59–63, 178–179, and 182 bp, respectively, before the *rbc*L protein coding regions. The putative promoter regions preceding the spinach and tobacco genes were found to be highly homologous and to contain prokaryote-like '–10' and '–35' regions (39), while the corresponding putative promoter from maize did not (28). In addition, the 5′ ends for the *atp*B genes from maize (20) and tobacco (38) were positioned 298–302 and 256–257 bp, respectively, before the *atp*B protein coding region. While each of the *atp*B putative promoter regions contained '–10' and '–35' regions, they did not share any extensive sequence homology with each other. The interesting possibility existed that these differences in putative promoter regions reflected different sequence requirements for plastid gene regulation in C3 and C4 plants. The results presented in this paper, however, show that transcription of these two genes is not as different as initially presumed.

In our reexamination of the 5′ termini of the *rbc*L and *atp*B transcripts we have chosen, whenever possible, to employ single-stranded DNAs as probes for S1 nuclease protection experiments. In addition, we have varied the RNA:DNA ratio to ensure that the DNA probe is present in excess with respect to the RNA. Under our experimental conditions, if a vast excess of RNA is used for S1 nuclease protection the largest transcript preferentially hybridizes to the DNA probe, and signals due to smaller transcripts are not detected. To avoid other potential artifacts of the S1 nuclease protection technique, we have also confirmed the position of 5′ termini by primer extension analysis. From these experiments we have found that multiple transcripts are common for the *rbc*L and *atp*B genes. In particular, the *rbc*L gene typically has two species of mRNAs, a '–65' transcript and a larger '–180' (spinach, pea, and tobacco) or '–300' (maize) transcript.

As a first step toward elucidating the regulatory mechanism of a particular chloroplast gene, it is essential to define the site of transcription initiation. The presence of two mRNA species for the *rbc*L gene could arise from separate transcription initiation events, as with certain operons in *E. coli* (30) and *B. subtilis* (17). Alternatively, these two transcripts could be the consequence of a single transcription initiation event followed by processing of the RNA, as with the cytochrome b gene in *S. cerevisiae* mitochondria (5). To differentiate between these possibilities, we have analyzed the different *rbc*L transcripts for the presence of 5′ terminal polyphosphates. By labeling chloroplast RNA with alpha-^{32}P-GTP and guanylyltransferase, we have demonstrated that transcription initiation of the pea and spinach *rbc*L genes occurs at the '–180' position. To our knowledge, this is the first characterization of an *in vivo* transcription initiation site for a higher plant plastid gene. While we did not observe 5′ end labeling of the maize *rbc*L transcripts, the DNA sequences that precede the *rbc*L '–180' (spinach, pea, and tobacco) and '–300' (maize) 5′ ends are highly conserved (Fig. 7). In addition, initiation of the pea, spinach, and maize *rbc*L genes occurs *in vitro* at these positions (Oroczo EM, Jr, Mullet JE, Hanley-Bowdoin L, Chua N-H, unpublished observations). For these reasons, we suggest that the transcription start site for the maize *rbc*L gene is located at position '–300'. In pea and spinach the *rbc*L '–65' transcript is not a substrate for guanylyltransferase; therefore, this transcript is probably derived from processing of a larger (e.g. '–180') mRNA.

We have also assayed the spinach, maize, and pea *atp*B mRNAs for the presence of 5′ terminal polyphosphates. Unfortunately, we were not able to detect 5′ end labeling of these transcripts by guanylyltransferase, presumably due to their low abundance relative to the *rbc*L transcripts. However, we note that the DNA sequence preceding the 5′ end of the largest *atp*B transcript is highly conserved (Fig. 10A). We have also observed production *in vitro* of the spinach '–455' transcript, the pea '–350' transcript, and the maize '–300' transcript with chloroplast RNA polymerase (Mullet JE, Orozco EM, Jr, Hanley-Bowdoin L, Chua N-H, unpub-

Table 1. Conservation of distance between the *rbc*L and *atp*B genes in higher plants.

Plant	*rbc*L 5′ Termini	*atp*B 5′ Termini	Intergenic distance
Spinacea oleracea	−178 to −179 ('−180') −64 ('−65')	−453 to −454 ('−455') −272 to −273 ('−275') −179 ('−180') −99 ('−100')	152 bp
Zea mays	−300 ('−300') −59 to −63 ('−65')	−298 to −302 ('−300')	157 bp
Pisum sativum	−178 ('−180') n.d. ('−65')	−351 to −355 ('−350')	154 bp
Nicotiana tabacum	−182 ('−180')	n.d. ('−490') −256 to −257	145 bp

The precise 5′ ends of *rbc*L and *atp*B transcripts as determined by S1 nuclease protection and/or primer extension analyses are shown. The approximate 5′ terminal positions of various transcripts (as referenced to in the text) are shown in parentheses. The corresponding *rbc*L-*atp*B intergenic distance of each plant is given. The 5′ ends of the '−180' *rbc*L transcripts from spinach (49), tobacco (39), and pea (G. Zurawski, personal communication) and the 5′ ends of the *atp*B transcripts from maize ('−300'; 20), pea ('−350'; G. Zurawski, personal communication) and tobacco ('−256' to '−257'; 38) have been determined previously. The spinach '−180' *atp*B and the pea '−65' *rbc*L transcripts are minor species (this report). The existence of a putative '−490' *atp*B transcript from tobacco and the corresponding *rbc*L-*atp*B intergenic distance are based on DNA sequence homologies as discussed in the text. The precise 5′ ends of the pea '−65' *rbc*L mRNA and the putative tobacco '−490' *atp*B mRNA were not determined.

lished observations). For these reasons we assign the transcription initiation site of the spinach *atp*B gene to the '−455' position.

The data presented in this paper indicate that the size of the intergenic region between the *rbc*L and *atp*B genes is conserved at approximately 150 bp among spinach, maize, and pea, and potentially for tobacco also (Table 1). Therefore, we have identified a relatively short region of DNA that should contain sequences important for plastid gene transcription. Within this intergenic region, the DNA sequence 5′ to the *rbc*L and *atp*B genes is highly conserved, and the sequence is divergent for the central region. We note that the conserved regions contain sequences reminiscent of the prokaryotic '−10' and '−35' promoter sequence, but their functional significance in chloroplast gene transcription remains to be established. In this connection we have recently developed a homologous *in vitro* system from higher plant chloroplasts that will correctly initiate transcription of these genes (Orozco, Mullet & Chua, manuscript in preparation). Experiments are in progress to determine what specific sequences within this 150 bp region are required for transcription initiation *in vitro*.

The question remains regarding the physiological significance of an RNA population with heterogeneity at the 5′ end. The functional significance of this heterogeneity in *rbc*L and *atp*B transcripts may be related to the translational properties of the various mRNAs. The large *rbc*L transcript is much more abundant in dicots (pea and spinach) than in monocots (maize). The barley *rbc*L transcripts have recently been characterized and are found to be similar to maize with respect to the DNA sequence and the presence of two 5′ ends at positions '−300' and '−65' (33, 51). In addition, the larger barley transcript appears to accumulate upon illumination of etiolated plants (33). In light-grown spinach the predominant *rbc*L transcript is the '−180' species, whereas in etiolated spinach there is a slight increase (approximately 5-fold) in the ratio of '−65':'−180' transcripts (Orozco EM, Jr, Mullet JE, Chua N-H, unpublished observations). This difference in the ratio of *rbc*L transcripts under varying physiological conditions will be an interesting subject for further study.

Acknowledgements

This work was supported by NIH grant GM 30726 and a grant from the Monsanto Company. J. E. M. was supported by NIH postdoctoral

fellowship 5F32GM08114. We are grateful to G. Lorimer for providing us with spinach grown in hydroponics, to H. Weissbach and G. Zurawski for communicating results prior to publication, and to R. E. Zielinski for discussions regarding S1 nuclease protection experiments. We also thank S. Zavodny and J. Carter for valuable technical assistance, J. O'Donnell for graphic illustrations and W. Roine for assembling this manuscript.

References

1. Barraclough R, Ellis RJ: The biosynthesis of ribulose bisphosphate carboxylase. Uncoupling of the synthesis of the large and small subunits in isolated soybean leaf cells. Eur J Biochem 94:165–177, 1979.
2. Bartlett SG, Grossman AR, Chua N-H: *In vitro* synthesis and uptake of cytoplasmically-synthesized chloroplast proteins. In: Edelman M, Hallick RB, Chua N-H (eds) Methods in Chloroplast Molecular Biology. Elsevier Biomedical Press, Amsterdam, The Netherlands, 1982, pp 1081–1091.
3. Boege F, Westhoff P, Zimmermann K, Zetsche K: Regulation of the synthesis of ribulose-1,5-bisphosphate carboxylase and its subunits in the flagellate *Chlorogonium elongatum*. I. The effect of light and acetate on the synthesis and the degradation of the enzyme. Eur J Biochem 113:581–586, 1981.
4. Broglie R, Coruzzi G, Keith B, Chua N-H: Molecular biology of C_4 photosynthesis in *Zea mays:* differential localization of proteins and mRNAs in the two leaf cell types. Plant Mol Biol 3:431–444, 1984.
5. Christianson T, Edwards JC, Mueller DM, Rabinowitz M: Identification of a single transcriptional initiation site for the glutamic tRNA and COB genes in yeast mitochondria. Proc Nat Acad Sci USA 80:5564–5568, 1983.
6. Cohen SN, Miller CA: Non-chromosomal antibiotic resistance in bacteria. II. Molecular nature of R-factors isolated from *Proteus mirabilis* and *Escherichia coli.* J Mol Biol 50:671–687, 1970.
7. Erion JL, Tarnowski J, Peacock S, Caldwell P, Redfield B, Brot N, Weissbach H: Synthesis of the large subunit of ribulose-1,5-bisphosphate carboxylase in an *in vitro* partially defined *E. coli* system. Plant Mol Biol 2:279–290, 1983.
8. Feierabend J, Wildner G: Formation of the small subunit in the absence of the large subunit of ribulose-1,5-bisphosphate carboxylase in 70S ribosome-deficient rye leaves. Arch Biochem Biophys 186:183–291, 1978.
9. Feierabend J, deBoer J: Comparative analysis of the action of cytokinin and light on the formation of ribulosebisphosphate carboxylase and plastid biogenesis. Planta 142:75–82, 1978.
10. Fluhr R, Fromm H, Edelman M: Clone bank of *Nicotiana tabacum* chloroplast DNA: mapping of the alpha, beta and epsilon subunits of the ATPase coupling factor, the large subunit of ribulose bisphosphate carboxylase, and the 32-kDal membrane protein. Gene 25:271–280, 1983.
11. Frankfort HM, Robertson HD: Ribonucleases. In: Linn SM, Roberts RJ (eds) Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982, pp 359–366.
12. Hallick RB, Richards OC, Gray PW: Isolation of intact, superhelical chloroplast DNA from *Euglena gracilis.* In: Edelman M, Hallick RB, Chua N-H (eds) Methods in Chloroplast Molecular Biology. Elsevier Biomedical Press, Amsterdam-New York-Oxford, 1982, pp 281–293.
13. Hallick RB, Bottomley W: Proposals for the naming of chloroplast genes. Plant Mol Biol Reporter 1:38–43, 1983.
14. Hansen U, Tenen DG, Livingston DM, Sharp PA: T antigen repression of SV40 early transcription from two promoters. Cell 27:603–612, 1981.
15. Huber SC, Hall TC, Edwards GE: Differential localization of fraction I protein between chloroplast types. Plant Physiol 57:730–733, 1976.
16. Iwanij V, Chua N-H, Siekevitz P: Synthesis and turnover of ribulose bisphosphate carboxylase and of its subunits during the cell cycle of *Chlamydomonas reinhardtii.* J Cell Biol 64:572–585, 1975.
17. Johnson WC, Moran CP, Losick R: Two polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. Nature 302:800–804, 1983.
18. Kawashima N, Wildman SG: Fraction I protein. Ann Rev Plant Physiol 21:325–358, 1970.
19. Kirchanski SJ, Park RB: Comparative studies of the thylakoid proteins of mesophyll and bundle-sheath plastids of *Zea mays.* Plant Physiol 58:345–349, 1976.
20. Krebbers ET, Larrinua IM, McIntosh L, Bogorad L: The maize chloroplast genes for the beta and epsilon subunits of the photosynthetic coupling factor CF_1 are fused. Nuc Acids Res 10:4985–5002, 1982.
21. Kung S: Tobacco fraction I protein: A unique genetic marker. Science 191:429–434, 1976.
22. Lee DC, Luse DS: Mapping the 5′ ends of *in vitro* synthesized RNAs with reverse transcriptase. Focus 4:1–3, 1982.
23. Levens D, Ticho B, Ackerman E, Rabinowitz M: Transcription initiation and 5′ termini of yeast mitochondrial RNA. J Biol Chem 256:5226–5232, 1981.
24. Link G, Coen DM, Bogorad L: Differential expression of the gene for the large subunit of ribulose bisphosphate carboxylase in maize leaf cell types. Cell 15:725–731, 1978.
25. Lizardi PM: Schleicher and Schuell publication #364, 1983.
26. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982.
27. Maxam AM, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds) Methods in Enzymology, Vol 65. Academic Press, New York, 1980, pp 499–560.
28. McIntosh L, Poulsen C, Bogorad L: Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase of maize. Nature 288:556–560, 1980.
29. Moss B: Utilization of the guanylyltransferase and methyltransferases of vaccinia virus to modify and identify the 5′-terminals of heterologous RNA species. Biochem Biophys Res Comm 74:374–383, 1977.
30. Musso RE, DiLauro R, Adhya S, deCrombrugghe B: Dual

control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. Cell 12:847–854, 1977.

31. Orozco EM Jr, Gray PW, Hallick RB: *Euglena gracilis* chloroplast ribosomal RNA transcription units. I. The location of transfer RNA, 5S, 16S, and 23S ribosomal RNA genes. J Biol Chem 255:10991–10996, 1980.
32. Palmer JD, Edwards H, Jorgensen RA, Thompson WF: Novel evolutionary variation in transcription and location of two chloroplast genes. Nuc Acids Res 10:6819–6832, 1982.
33. Poulsen C: Two mRNA species by 258 nucleotides at the 5′ end are formed from the barley chloroplast *rbc*L gene. Carlsberg Res Comm 49:89–104.
34. Rosenberg M, Court D: Regulatory sequences involved in the promotion and termination of RNA transcription. Ann Rev Genet 13:319–353, 1979.
35. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Nat Acad Sci USA 74:5463–5476, 1977.
36. Sasaki Y, Ishiye M, Sakihama T, Kamikubo T: Light-induced increase of mRNA activity coding for the small subunit of ribulose-1,5-bisphosphate carboxylase. J Biol Chem 256:2315–2320, 1981.
37. Shenk T: Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. In: Shatkin AJ (ed) Initiation Signals in Viral Gene Expression. Springer Verlag, Berlin-Heidelberg-New York, 1981, pp 25–46.
38. Shinozaki K, Sugiura M: Sequence of the intercistronic region between the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit and the coupling factor beta subunit gene. Nuc Acids Res 10:4923–4934, 1982.
39. Shinozaki K, Sugiura M: The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase. Gene 20:91–102, 1982.
40. Shinozaki K, Sasaki Y, Sakihama T, Kamikabo T: Coordinate light-induction of two mRNAs, encoded in nuclei and chloroplasts, of ribulose-1,5-bisphosphate carboxylase/oxygenase. FEBS Lett 144:73–76, 1982.
41. Shinozaki K, Deno H, Kato A, Sugiura M: Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase. Gene 24:147–155, 1983.
42. Smith SM, Ellis RJ: Light-stimulated accumulation of transcripts of nuclear and chloroplast genes for ribulose bisphosphate carboxylase. J Mol Appl Genet 1:127–137, 1981.
43. Sollner-Webb B, Reeder RH: The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell 18:485–499.
44. Tjian R: Regulation of viral transcription and DNA replication by the SV40 large T antigen. In: Shatkin AJ (ed) Initiation Signals in Viral Gene Expression. Springer-Verlag, Berlin-Heidelberg-New York, 1981, pp 5–24.
45. Tobin E, Suttie JL: Light effects on the synthesis of ribulose-1,5-bisphosphate carboxylase in *Lemna gibba*. Plant Physiol 65:641–647, 1980.
46. Viera J, Messing J: The pUC plasmids, an m13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268, 1982.
47. Weaver RF, Weissmann C: Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of the 15S beta-globin mRNA precursor and mature 10S beta-globin mRNA have identical map coordinates. Nuc Acids Res 7:1175–1193, 1979.
48. Whitfeld PR, Bottomley W: Organization and structure of chloroplast genes. Ann Rev Plant Physiol 34:279–310, 1983.
49. Zurawski G, Perrot B, Bottomley W, Whitfeld PR: The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase form spinach chloroplast DNA. Nuc Acids Res 9:3251–3270, 1981.
50. Zuraswki G, Bottomley W, Whitfeld PR: Structure of the genes for the beta and epsilon subunits of spinach chloroplast ATPase indicate a dicistronic mRNA and an overlapping translation stop/start signal. Proc Nat Acad Sci USA 79:6260–6264, 1982.
51. Zurawski G, Clegg MT, Brown AHD: The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735–749, 1984.

Received 5 July 1984; in revised form 23 July 1984; accepted 30 July 1984.