Plant Molecular Biology 6: 229–243, 1986

Characterization of *P. sativum* chloroplast psbA transcripts produced *in vivo*, *in vitro* and in *E. coli*

Scott K. Boyer & John E. Mullet[1]
Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843, U.S.A.

Keywords: chloroplast transcription, *E. coli* transcription, psbA, nucleotide sequence comparison, RNA mapping

Summary

We have analyzed the region of the chloroplast genome from *P. sativum* which encodes the 5′-end of psbA. S_1 nuclease mapping and primer extension analysis of chloroplast RNA revealed psbA transcripts with 5′-termini 92, 93 and 68 nucleotides upstream from the psbA open reading frame. The psbA transcripts with 5′-ends 92–93 nucleotides upstream from the psbA open reading frame can be labeled with alpha-^{32}P-GTP by guanylyltransferase. DNA sequences 10 and 35 bp upstream from the longest psbA transcript showed homology to −10 and −35 consensus promoter sequences in *E. coli*. Truncated psbA constructs which contain the putative psbA promoter sequences were shown to promote transcription in *E. coli* from a site similar to that used by chloroplast RNA polymerase *in vivo*. Accurate transcription of psbA constructs was also observed in a homologous *in vitro* transcription extract from chloroplasts. Sequence analysis of the region upstream from the psbA transcripts revealed a putative 3′-exon of a tRNA-Lys (240 bp upstream) and an unidentified open reading frame (URF, 485 bp upstream). The 3′-end of the URF mRNA was located approximately 240 bp from the 5′-end of the longest psbA transcript indicating that the URF and tRNA-Lys sequence are cotranscribed. Comparison of *P. sativum* and *N. tabacum* DNA sequences at the 5′-end of psbA revealed homology between sequences coding for psbA mRNA including 40 bp upstream from the longest psbA transcript. A second region of homology which includes the tRNA-Lys sequence was also located. In contrast the intergenic DNA exhibited extensive divergence in size and sequence.

Introduction

Chloroplast DNA of the higher plant *Pisum sativum* is approximately 120 kbp in size (33) and codes for tRNAs (6), rRNAs and mRNAs (42). The DNA sequences which are important for the transcription of various genes coded by chloroplast DNA have been assessed by analysis of DNA sequences surrounding the 5′-ends of transcripts and through the use of *in vitro* transcription systems (11, 12, 14, 19, 28, 32). In the case of protein coding genes it has been shown that accurate transcription of chloroplast templates can be achieved *in vitro* (19, 32). Analysis of DNA sequences required for accurate transcription of chloroplast templates *in vitro* have confirmed the importance of a sequence located 35 bp upstream from the site of transcription initiation (19). This sequence corresponds to the '−35' consensus sequence which is involved in prokaryotic gene transcription initiation (5).

The 5′-ends of several chloroplast protein gene mRNAs have been found to be heterogeneous. For example, two 5′-ends have been reported for rbcL transcripts (7, 28) and one to three 5′-ends for atpB mRNA (28). In these cases the location of the 5′-end of the longest transcript was also a site of transcription initiation (28). Available data indicates that the shorter transcripts for rbcL are de-

[1]Person to whom correspondence should be addressed; 409-845-0722.

rived from the primary transcript by RNA processing (28). These results show that a 5′-end of a chloroplast transcript does not necessarily correspond to a site of transcription initiation and that further tests are required to establish this point. The function of 5′-end mRNA heterogeneity for genes such as rbcL and atpB is unknown although it has been suggested that different lengths of untranslated mRNA may be involved in regulation of translation (28).

Gene expression in plastids is reported to be regulated at the transcriptional level during plastid biogenesis by light (35). This was first reported for the psbA gene in maize (3, 29) and has also been observed in mustard (18). In maize, psbA represents one of a class of genes whose mRNA levels are induced by exposure of dark-grown plants to light (35). In mustard, however, psbA is at present the only example of light mediated induction (20). In other plants such as peas, mung bean, and spinach some psbA mRNA is found in dark-grown plants but during light-induced development psbA mRNA levels increase (15, 40). In contrast, in mature Spirodela psbA transcript levels are quite stable although translation of protein from these transcripts varies considerably (10).

The psbA gene product is a 32 kDa membrane polypeptide which is an integral component of the photosystem II-chlorophyll-protein complex (24) and is a membrane-associated chloroplast protein (4). The 32 kDa polypeptide is a quinone-binding component in photosystem II and is the site of action of a class of photosynthetic herbicides (24, 34, 37). The DNA sequence of psbA has been reported for several plants (16, 17, 21, 31, 36, 38, 44). From the DNA sequence a 38500 dalton primary translation product is predicted with almost the entire deduced amino acid sequence conserved between the species examined (16, 31, 38, 44). In this paper we have focused our study on the function of the sequences between the psbA open reading frame and the gene whose 3′-end is located upstream from psbA. The transcription of psbA in *E. coli* and in a homologous *in vitro* transcription system is reported here.

Materials and methods

Reagents

Ribonucleotides, deoxyribonucleotides and dideoxyribonucleotides were from Sigma Chemical Co. or P-L Biochemicals, Inc. ^{32}P-nucleotides were obtained from New England Nuclear. Trace proteins were removed from the *E. coli* tRNA type XXI (Sigma Chemical Co.) used as carrier nucleic acid by phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol) extraction and the tRNA was stored in H_2O at −20 °C. ∅X174-HaeIII DNA fragments were from Bethesda Research Laboratories, Inc. Electrophoresis grade 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 analytical grade mixed bed resin (Bio-Rad Laboratories) and stored at −20 °C. Analytical and preparative polyacrylamide gels were prepared with acrylamide and electrophoresis grade N,N′-methylenebisacrylamide (Eastman Kodak Co.) that was treated with activated charcoal (Fisher Scientific Co.) for one hour and twice filtered before use. NA-45 membranes for nucleic acid isolation were from Schleicher and Schuell, Inc.

Enzymes

Restriction endonucleases, T_4 polynucleotide kinase, T_4 DNA ligase, and the large fragment of DNA Polymerase I were from BRL, Inc. and New England Biolabs, Inc. Calf intestine alkaline phosphatase was from Boehringer-Mannheim. S_1 nuclease, AMV reverse transcriptase and guanylyltransferase were from BRL, Inc. DNase I was from Cooper Biomedicals.

Plant growth conditions

Pisum sativum (Progress No. 9) was purchased from Burpee Seed Co. Plants were grown in vermiculite in a Conviron environmental growth chamber in continuous light at 23 °C. Plants were harvested 7 days after inhibition of dry seeds.

Chloroplast nucleic acid isolation

Intact chloroplasts were isolated from percoll

gradients by the method of Bartlett *et al.* (2). Pea chloroplast DNA was isolated from plastids by centrifugation through CsCl-ethidium bromide gradients essentially as described by Hallick *et al.* (13). Chloroplast RNA was obtained by lysis of intact plastids with an equal volume of RNA extraction buffer (0.3 M NaCl, 1% sodium dodecylsulfate, 20 mM EDTA, 10 mM Tris-HCl, pH 8) and subsequently extracted 2 times with phenol:chloroform: isoamyl alcohol (25:24:1; vol:vol:vol). The aqueous phase was then precipitated and stored at −80 °C in TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

E. coli *RNA Isolation*

An overnight culture of the appropriate plasmid containing *E. coli* strain was diluted into 5 ml of media and grown to mid-log phase. The culture was poured into a prechilled, phenol-rinsed mortar. Two volumes of liquid N_2 was used to quick-freeze the cells and a pestle was used to grind the cells to a powder. Before the cells thawed, 10 ml of RNA extraction buffer and 15 ml of phenol:chloroform: isoamyl alcohol (25:24:1; vol:vol:vol) were added and the solution was stirred until homogeneous. The solution was then transferred to a chilled 30 ml corex tube (siliconized and heat treated), vortexed and then centrifuged at $5000 \times g$ in a swinging bucket type rotor. The aqueous phase was ethanol precipitated in corex tubes, rinsed with 80% ethanol and the pellet dried. The pellet was resuspended in 0.3 ml of TE and transferred to a 1.5 ml eppendorf tube. Undissolved material was removed by centrifugation in an eppendorf centrifuge for 10 min at 4 °C. The supernatant was ethanol precipitated and the pellet resuspended in 0.1 ml of TE. The nucleic acid solution was then stored at −80 °C and was later used for S_1 nuclease assays.

Isolation of DNA restriction fragments

After the DNAs were digested with the appropriate restriction enzymes the fragments were separated electrophoretically through agarose or polyacrylamide gels. Restriction fragments were isolated by electroelution into dialysis bags (23) or by electrophoresis onto a NA−45 membrane (22).

Plasmid constructions

DNA fragments which contain the 5′ or 3′-end of psbA from pea were identified by Palmer & Thompson (33). The 709 bp XbaI-PstI fragment from pea chloroplast DNA was inserted into pUC18 and pUC19 (26, 30) to generated pPPBX-10218 and pPPBX-10219, respectively. Plasmids transformed into the bacterial strain TB1 (1) were isolated from CsCl-ethidium bromide gradients (23).

Sequencing and preparation of ^{32}P-DNAs

The 1308 bp BamHi-PstI fragment was sequenced by the M13/dideoxy sequencing method (26). Linear DNAs or synthetic oligonucleotide were 5′-end labeled with gamma ^{32}P-ATP and T4 polynucleotide kinase (23) or 3′-end labeled with the appropriate alpha ^{32}P-nucleotide triphosphate (dNTP) and the large fragment of DNA Polymerase I (26). The ^{32}P-DNAs were separated on 5 or 8% polyacrylamide gels. After electrophoresis of the ^{32}P-DNAs on acrylamide gels, the separated fragments were purified by electroelution into dialysis bags (23). In addition to sequencing by the M13/dideoxy method of Messing (26), the 117 bp MboI-HinfI (HinfI labeled) was sequenced by the method of Maxam & Gilbert (25).

S_1 nuclease protection assay

Excess end-labeled ^{32}P-DNA (5000 to 20000 cpm/assay) and carrier *E. coli* tRNA (10 μg/assay) were resuspended in 80% formamide (10 μl/assay) and boiled for 5 min. The ^{32}P-DNA-formamide solution was added to a dry pellet of S_1 hybridization buffer (final concentration: 400 mM NaCl, 40 mM Pipes-NaOH pH 6.4, 1 mM EDTA) and either chloroplast or *E. coli* RNA. All hybridizations were performed at 37 °C for 12−18 hours except for the fine mapping of *in vivo* psbA RNA which was done using the 117 bp MboI-HinfI fragment which was slow cooled to room temperature. After hybridization the solution was diluted with 100 μl of S_1 assay buffer (250 mM NaCl, 30 mM NaOAc, pH 4.6, 1 mM $ZnSO_4$, 20 μg/ml denatured salmon sperm DNA), 10−30 units of S_1 nuclease was added and the reaction incubated at 30 °C for 40 min. The S_1-protected hybrid was

ethanol precipitated with 10 μg of carrier *E. coli* tRNA and analyzed on 8% polyacrylamide-8.3 M urea sequencing gels (25).

Primer extension assay

Double-stranded 5′-end labeled ^{32}P-DNA fragments were denatured and hybridized to chloroplast RNA as described in the S_1 nuclease protection assays. The hybrid nucleic acid was ethanol precipitated and resuspended in 20 μl of 50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, and 1 mM of each dNTP. AMV reverse transcriptase (25 units) was added and the reaction carried out at 37 °C. After the reaction and the mixture was ethanol precipitated with 10 μg *E. coli* tRNA, resuspended in 100 μl of S_1 nuclease assay buffer, and ethanol precipitated a second time.

The 5′ end-labeled synthetic fifteen nucleotide oligomer (see Fig. 2) was hybridized to chloroplast RNA by heating to 65 °C and slow cooling to 37 °C. The hybridization solution consisted of chloroplast nucleic acid (1.2 or 0.012 μg), 100000 cpm of ^{32}P-15-mer, 50 mM Tris-HCl (pH 8.2), 10 mM $MgCl_2$ and 50 mM KCl. After the mix had reached 37 °C, the four unlabeled dNTPs (1 mM each), 10 mM DTT and 25 units of AMV reverse transcriptase was added and the mix incubated at 37 °C for 60 min. Upon completion of the reaction 10 μg of *E. coli* tRNA were added and the mix was precipitated with ethanol. The ^{32}P-DNAs were then analyzed on 8% polyacrylamide-8.3 M urea sequencing gels (25).

Analysis of RNA by use of guanylyltransferase

Guanylyltransferase and alpha-^{32}P-GTP were used to 5′-end label pea chloroplast RNA as described by Mullet *et al.* (28). The 20-μl reaction mix was composed of 3–20 μg of chloroplast RNA, 10 units of vaccinia virus guanylyltransferase, 0.01 mCi of alpha-^{32}P-GTP, 50 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$ and 1 mM dithiothreitol and was incubated at 37 °C for 30 min. The reaction was stopped by adding 80 μl of RNA extraction buffer and extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol). Carrier tRNAs (10 μg) were added and the mixture was twice precipitated with ethanol. Unlabeled double- or single-stranded DNAs (100 to 200 pmoles/assay) were hybridized to the chloroplast ^{32}P-RNA (50000 cpm/assay) and treated with S_1 nuclease as previously described.

Preparation of transcription extracts from pea

Intact chloroplasts were purified as previously described (2). The plastids were lysed and extracts were prepared through ammonium sulfate precipitation as described by Orozco *et al.* (32). After centrifugation, the resulting pellet was resuspended in a minimal volume 'DE-52 buffer' plus glycerol [25 mM Tricine KOH, pH 8.0, 50 mM KCl, 0.1 mM EDTA, 0.5 mM DTT, 0.05 mM epsilon-amino-caproic acid (Sigma Chemical Co.), 0.4 mM phenylmethylsulfonyl-fluoride (Sigma Chemical Co.), 5% glycerol] and was dialyzed 12 h in DE-52 buffer plus glycerol. The resulting 'transcription extract', equivalent to the 'high salt extract' of Orozco *et al.* (32), was divided into aliquots, quick-frozen in liquid N_2, and stored at −80 °C.

Proteinase K treatment of DNase I

Since it has been shown that calcium ions cause DNase I to become resistant to proteinase K digestion (41), RNases were removed from a DNase I solution by treatment with proteinase K as described by Orozco *et al.* (32). The DNase I-proteinase K mixture was used immediately after preparation.

In vitro *transcription assays*

In vitro transcription assays were performed essentially as described by Orozco *et al.* (32). The 50-μl *in vitro* transcription reaction mixture consisted of DE-52 buffer with 0.5 mM CTP, 0.5 mM GTP, 0.05 mM ATP and 0.05 mM UTP. In addition to the previously mentioned components, the reaction mixture contain variable amounts of 'transcription extract' and 'template' (purified plasmid DNAs) to optimize production of *in vitro* RNA. The reaction was incubated at 26 °C for 30 min and was followed by addition of 25 μg of carrier tRNA, DNase I buffer (to final concentration of 20 mM Tris HCl, pH 8.0, 10 mM $CaCl_2$) and proteinase K treated-DNaseI (25 μg DNase I/assay). The addition of *E. coli* tRNA increased the amount of RNA detected in S_1 protection assays by approximately

10-fold. The mixture was treated with DNaseI for 15 min at 37 °C and the reaction stopped by the addition of TE to 150 µl and subsequent addition of 150 µl of RNA extraction buffer. Phenol:chloroform:isoamyl alcohol extraction was performed after the addition of 10 µg of carrier tRNA. The *in vitro* RNA was precipitated twice with ethanol and then analyzed by S_1 nuclease protection assays as previously described.

Results

DNA sequence analysis of the psbA 5'-region from P. sativum

The restriction map for a 2.6 kbp BamHI-EcoRI segment of the pea ctDNA which contains the 5'-end of psbA is shown in Fig. 1. Restriction sites used for the preparation of fragments for S_1 nuclease protection, primer extension, and DNA sequencing experiments are also indicated in this figure. The DNA sequence for the 1308 bp BamHI-PstI restriction fragment was determined by the M13/dideoxy method (26) (see Fig. 1c for sequence strategy; Fig. 2 for sequence). A comparison of our sequence, starting at position 795 with that of the recently published psbA sequences from *P. sativum* (31) showed differences at positions 799, 807, and 841 (see Fig. 2). A guanosine residue at position 799 is part of the recognition site for HphI, whereas the thymidine reported by Oishi *et al.* (31) would eliminate this restriction site. Restriction analysis with the HphI enzyme confirmed the presence of the guanosine at position 799 in our sequence. At position 807 we found a cytosine instead of the guanosine reported by Oishi *et al.* (31). The presence of a guanosine at this position would generate a HpaII recognition site. When this segment was analysized with HpaII or MspI (to eliminate negative results due to methylation) no restriction sites were found at this position.

Location of a transcript 3'-end upstream from the psbA gene

The sequence data in Fig. 2 revealed an open reading frame homologous to psbA and upstream

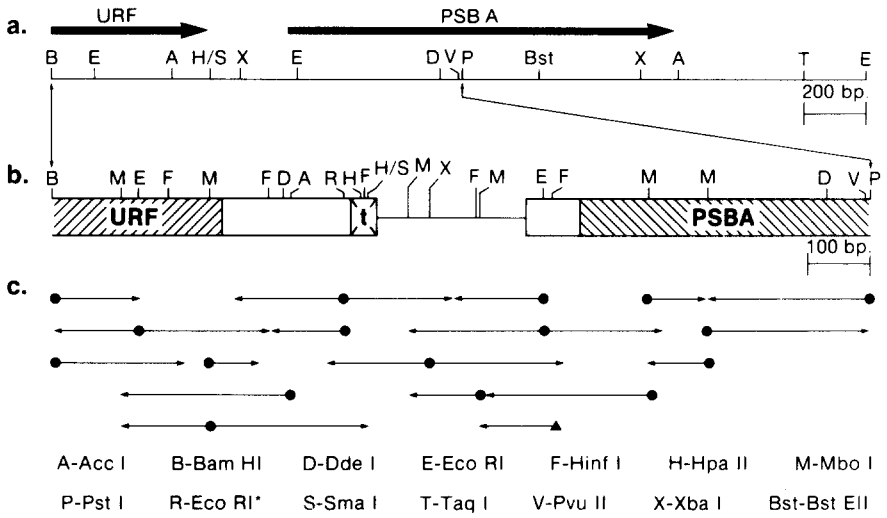


Fig. 1. Restriction map, transcript map, and sequencing strategy of the psbA-URF region from *P. sativum.* (a) The restriction map shows the 2.6 kbp BamHI-EcoRI restriction fragment and the location and direction of the protein coding mRNAs (solid arrows above the map) within this region. (b) The restriction map is of the 1308 bp BamHI-PstI subfragment of the 2.6 kbp BamHI-EcoRI fragment. The boxed regions are mRNA coding segments (of the largest RNA species for psbA) whereas the single line is the intergenic spacer region. The cross-hatched boxes indicate the open reading frames of the respective genes and the 't' indicates the putative 3'-exon of a gene coding tRNA-Lys. Restriction enzyme cleavage sites used in cloning, sequencing and probe preparation are indicated above this figure. (c) The sequencing strategy for the 1308 p BamHI-PstI fragment is shown at the bottom of this figure. Arrows originating from solid circles indicate sequencing by the method of Messing (26) and the arrow originating from the solid triangle indicates sequencing by the method of Maxam & Gilbert (25). The DNA sequence is shown in Fig. 2.

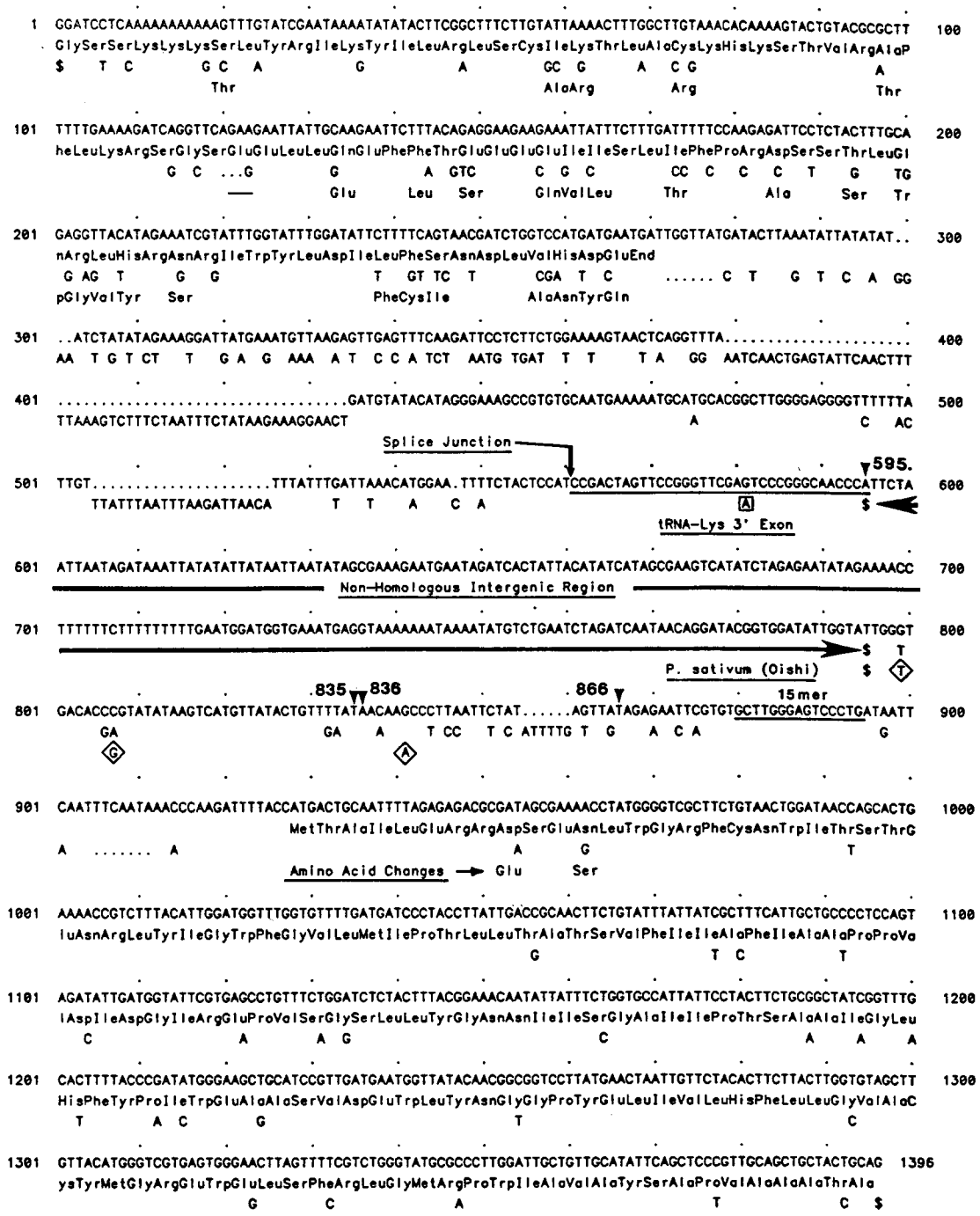
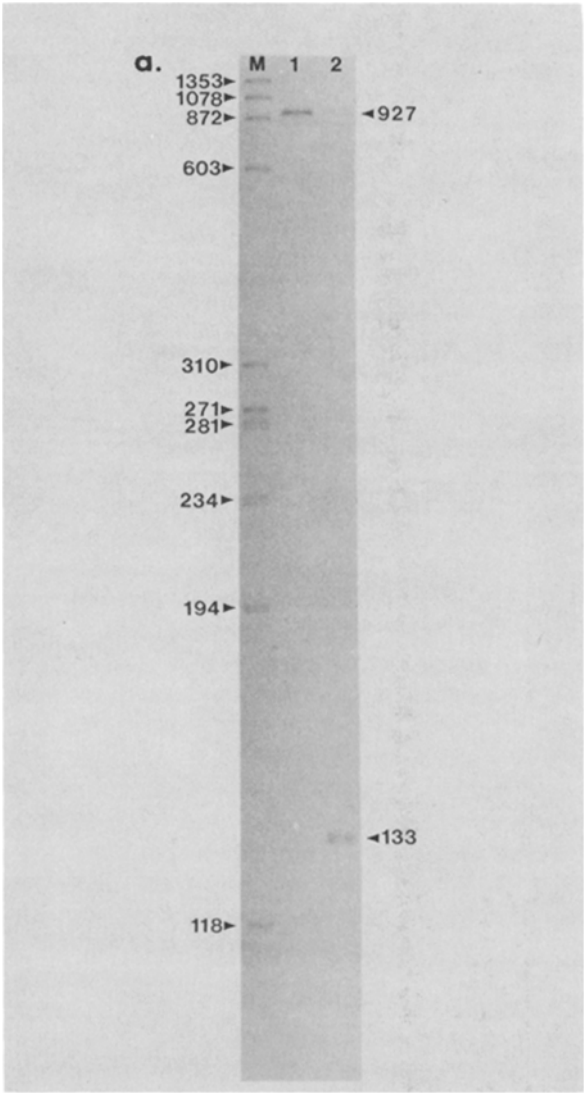


Fig. 2. The DNA sequence of the 1308 bp BamHI-PstI fragment from *P. sativum*. Line one (position 1 to 1396) is the primary DNA sequence of the 1308 bp fragment with gaps (...) introduced in the sequence to align the sequence with *N. tabacum* sequence. The second line from positions 1 through 270 shows the deduced amino acid sequence for the carboxyl terminus of the URF. The $ at position 1 and 595 indicate the regions of DNA homology (between *P. sativum* and *N. tabacum*) at the 3′ end of the URF, with only the differences with the *N. tabacum* sequences shown on line 3 (1 to 300) and line 2 (301 to 595). The tRNA-Lys 3′ exon is underlined between position 561 and 595 with the one nucleotide change with *N. tabacum* shown boxed below the exon. The non-homologous intergenic spacer region is shown between position 596 and 794. The comparision of nucleotide sequence is between the $ at position 795 and 1395, respectively, and differences with the *N. tabacum* are shown on lines two (from position 795 to 900) and three (from position 901 to 1396). The three nucleotide differences from the sequence published by Oishi *et al.* (31) are shown in diamonds on line three (between positions 795 and 900). The deduced amino acid sequence for *P. sativum* starts at position 928 on line 2 and the two amino acid changes with *N. tabacum* are shown on line 4. The procaryotic-like -35 and -10 promoter sequences are underlined and are centered at positions 802 and 826, respectively. The 15 nuc oligomer (used in the fine mapping of the 5′ end of the psbA transcripts(hybridizes to the underlined sequence from position 880 to 894 inclusive.

of this sequence a putative 3′-exon of a tRNA-Lys and an unidentified open reading frame of 89 amino acids (Figs. 1, 2). To determine if transcripts from the unidentified open reading frame and tRNA-Lys sequence terminated within the sequences 5′ to psbA a 927 bp AccI-PstI fragment (3′-end labeled at the AccI site) was used for S_1 nuclease protection assays (Fig. 3a). This assay revealed the presence of a 3′-terminus 133 nucleotides from the AccI site. A 3′-terminus was also mapped to the same region using a BamHI-PvuII DNA probe (data not shown). These data are consistent with cotranscription of the unidentified open reading frame and putative 3′-exon of tRNA-Lys.



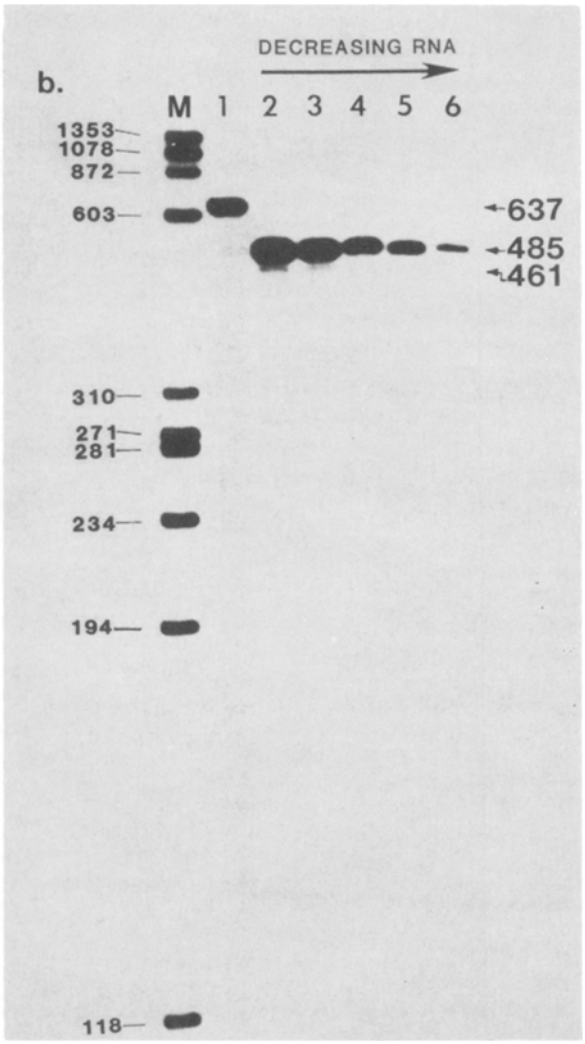
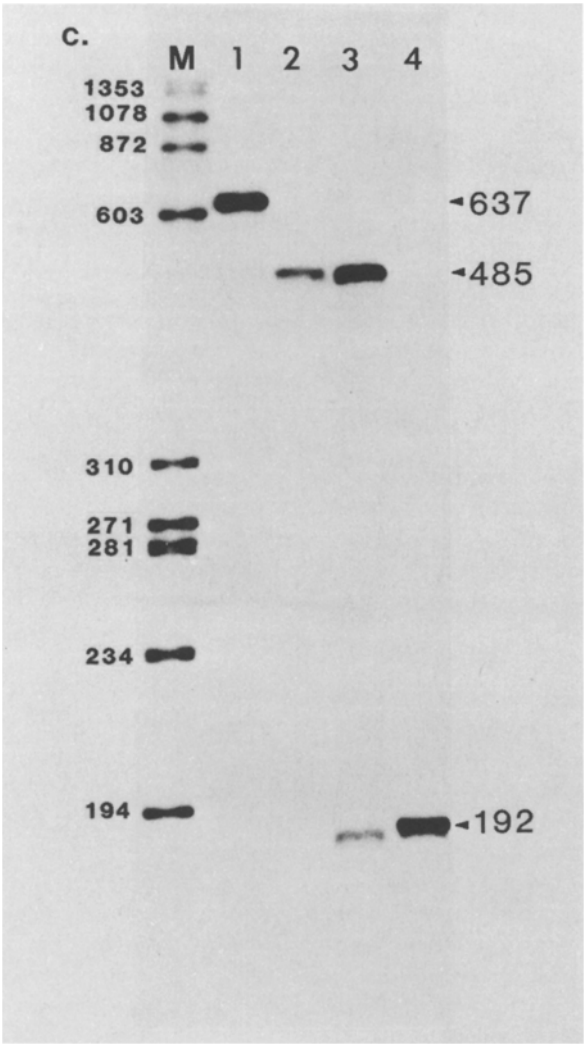


Fig. 3. a. The 3′-terminus of the URF transcript. Lane 1 shows the 927 nuc AccI-PstI probe (3′-end labeled at the AccI site). Lane 2 shows the S_1 nuclease protection experiment with a 133 nuc S_1 nuclease protected band. The molecular weights of the standards in Lane M are shown to the left of their respective bands.

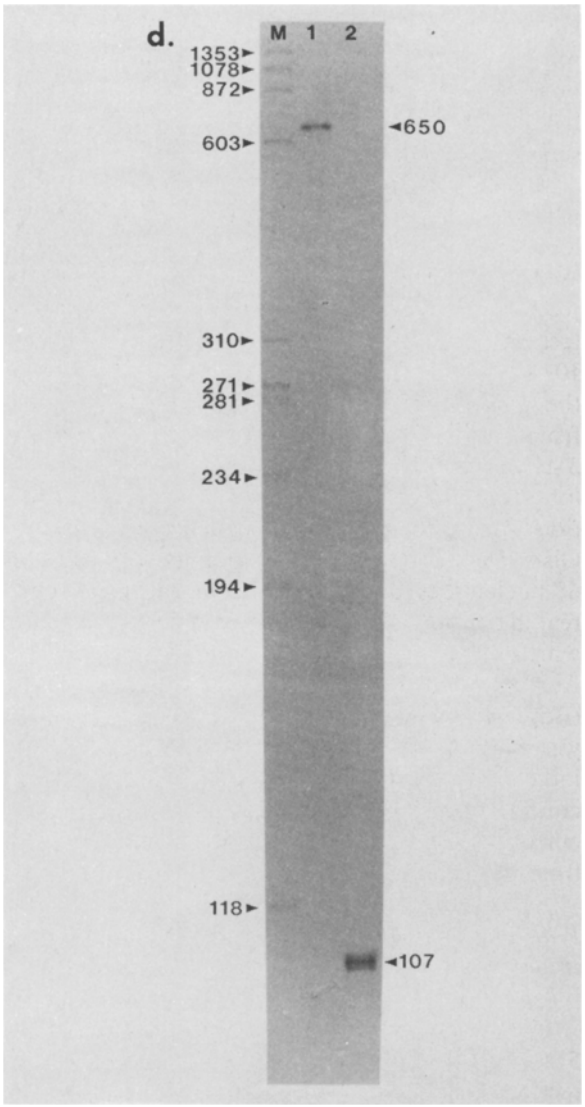
b. An S_1 nuclease protection experiment where the RNA concentration in the reaction mixture was varied. Lane 1 shows the 637 nuc XbaI-DdeI (5′-end labeled at the DdeI site) probe. Lanes 2 through 6 show a titration of total nucleic acid corresponding to 2.1, 0.7, 0.21, 0.07 and 0.021 μg of total chloroplast nucleic acid. The 485 nuc protected fragment was observed in all lanes and at the higher concentrations (lanes 2 and 3) the 461 nuc protected fragment is visible. The molecular weight markers are in lane M.

Localization of psbA RNA 5′ and 3′-termini

The 5′-terminus of the psbA gene was previously localized to a 5.0 kbp PstI fragment from the *P. sativum* chloroplast genome (33). Further analysis indicated that the 5′-end of the gene was located in a BamHI-PstI DNA fragment of the original PstI DNA region. A 637 bp XbaI-DdeI restriction fragment, 5′-end-labeled at the DdeI site, was used as a probe to determine the location and orientation of the psbA gene by S_1 nuclease protection experiments. The assay yielded two protected fragments of 461 nuc and 485 nuc, respectively (Fig. 3b). The relative amounts of each transcript was determined



c. The 5′-end analysis of pea psbA mRNA by S_1 nuclease assay and primer extension assay. The molecular weight markers are in lane M. Lane 1 is the 637 nucleotide XbaI-DdeI probe. Lane 2 is the S_1 nuclease protection experiment using 0.021 μg of total nucleic acid and shows a 485 nucleotide protected fragment. The primer extension assay is shown in Lane 3 and the 485 nucleotide extended primer is shown co-migrating with the fragment protected in the S_1 nuclease assays. Lane 4 shows the 192 nuc MboI-DdeI (5′-end labeled at the DdeI site) primer used for the primer extension assay.



d. The 3′-end analysis of the psbA gene from peas by S_1 nuclease protection assays. Lane 1 shows the 650 nuc XbaI-EcoRI (3′-end labeled at the XbaI site) probe and Lane 2 shows the S_1 nuclease protection experiment using the 650 nuc probe and 1.2 μg of total nucleic acid. A major band appears at approximately 107 nuc after digestion with S_1 nuclease. Lane M is the molecular weight standard ∅X174-HaeIII digested and 5′-end labeled.

by an S_1 titration where the amount of ctRNA was varied over three orders of magnitude. The shorter psbA transcript could only be seen at high RNA concentrations and always represented a minor percent of the total psbA transcripts.

In order to eliminate the possibility that the 5′-ends observed were S_1 nuclease artifacts, primer extension assays were conducted using a 5′-end-labeled MboI-DdeI (labeled at the DdeI end) fragment as a primer. At high concentrations of ctRNA both 5′-ends were observed (data not shown). At lower RNA concentrations only the larger transcript's 5′-end was observed (Fig. 3c).

A 650 bp XbaI-EcoRI DNA fragment, 3′-end-labeled at the XbaI site was used as a probe for the localization of the 3′-termini of the psbA gene. When this probe was hybridized to ctRNA and subjected to S_1 nuclease, only one fragment was protected. The S_1 protected fragment is approximately 107 nucleotides in length (Fig. 3d). The location of the 3′-terminus for the psbA transcript derived from our S_1 protection experiment and the 3′-terminus determined by Oishi *et al.* (31) using R-loop mapping are different by more than 65 nucleotides. Using the nucleotide sequence published by Oishi *et al.* (31) the 3′-terminus of psbA is 68 nucleotides downstream from the psbA open reading frame. We also located a DNA sequence which codes for tRNA-His approximately 196 bp from the 3′-end of the psbA open reading frame (Boyer & Mullet, unpublished). Therefore our analysis of the 3′-terminius of psbA transcripts *in vivo* indicates that psbA and the tRNA-His gene are not contranscribed or that RNA processing has occurred to remove the tRNA-His from the psbA transcripts.

Fine mapping of the in vivo *5′-termini of psbA transcripts*

In order to determine the exact position of the 5′-termini of the psbA transcripts, S_1 nuclease protection assays were electrophoresed beside the 'Maxam & Gilbert' sequencing reactions for the S_1 probe (data not shown). An S_1 nuclease protection assay for the major transcript using the 119 bp MboI-HinfI 5′-end-labeled fragment (HinfI labeled) as the probe showed several bands with the most intense being a cytosine at position 838. In a similar experiment the position of the minor tran-

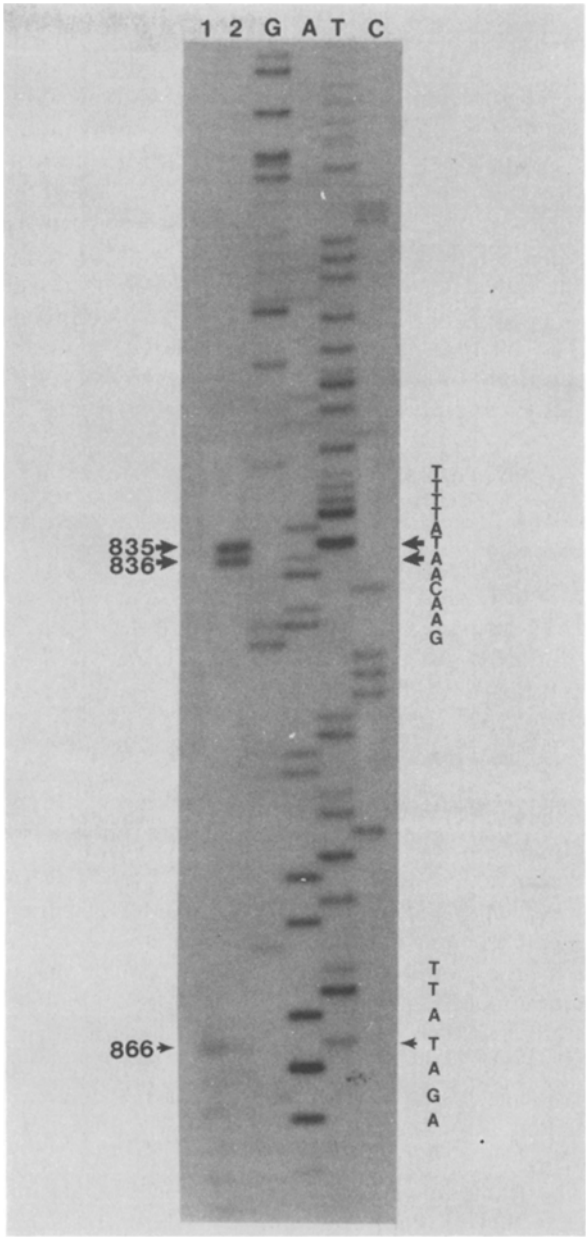


Fig. 4. Fine mapping of the psbA termini. The 5′-termini of psbA were verified by primer extension of the psbA mRNA using the (5′-end labeled) synthetic 15 nuc primer shown in Fig. 2. Lanes 1 and 2 show the primer extension assays of the psbA mRNA (1.2 and 0.012 μg total nucleic acid, respectively). This lane shows two intense bands at the adenosine residue corresponding to position 836 and at the thymidine residue at position 835. The minor RNA species noted at higher RNA concentrations (see Fig. 3b) whose 5′-end is the thymidine residue at position 866 is also shown in lane R. The dideoxy sequencing reactions lanes G A T and C, using the same primer are labeled as their complement so that the mRNA sequence can be read directly from the DNA sequence.

script was mapped to an thymidine at position 866.

Primer extension experiments using a 15 nuc synthetic oligomer as the primer showed three bands (Fig. 4). Major bands are seen at a thymidine residue at position 835 and an adenosine at position 836. These two bands mark the location of the major psbA 5′-terminus. This indicates that the other bands observed with the S_1 nuclease assay (described above) were artifacts which could be due to the highly AT rich sequence at the 5′-terminus. The position of a second minor RNA terminus at position 866 observed with S_1 assays was also verified by the primer extension assay (Fig. 4).

Identification of the transcription initiation site by 'capping'

The primary transcription product of a gene has a 5′-terminal triphosphate, whereas 5′-termini derived through RNA processing would have either a monophosphate or a hydroxyl group (9). This information can be used to verify a transcription initiation site since guanylyltransferase will only add GMP to the 5′-terminus of RNA containing a 5′-terminal polyphosphate (27).

Pea ctRNA was radioactively labeled with alpha-^{32}P-GTP using guanylyltransferase as described in Materials and Methods. The ^{32}P-RNA was then hybridized to a 1308 bp BamHI-PstI fragment or to single-stranded M13 viral DNA containing the same 1308 nuc insert. Equal molar amounts of each probe were hybridized to the ^{32}P-RNA and then subjected to S_1 nuclease treatment. When the resulting assays were electrophoresed on sequencing gels a single intense band of approximately the size expected for the largest psbA transcript was observed (Fig. 5, marked with an arrow). The single intense band showed slower migration than that of a corresponding ^{32}P-DNA which had been protected by psbA transcripts. Similar mobility differences have been previously noted by Mullet *et al.* (28) for rbcL transcripts. The apparent difference in mobility is probably due to the 'CAP-structure' and the structural differences between RNA and DNA. The five minor bands of lower molecular weight were found to be S_1 nuclease resistant even in the absence of any externally added probe. These minor bands were eliminated by pretreatment of the ^{32}P-RNA with DNase I (data not shown) indicating that the minor bands

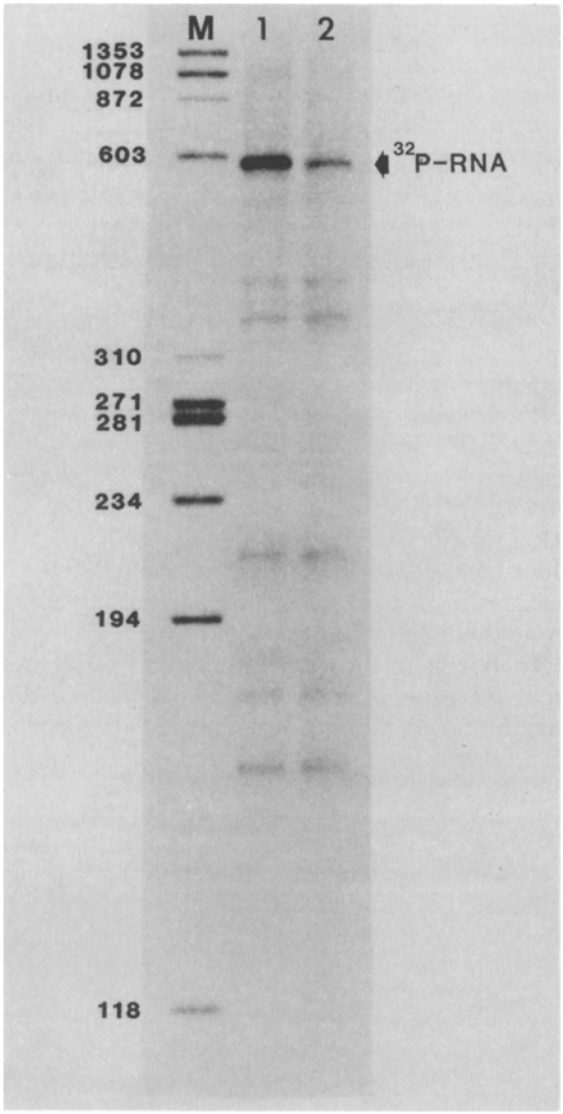


Fig. 5. S_1 nuclease protection assay of 'Capped' ctRNA reveals the protection of a major 590 nuc RNA. Lanes 1 and 2 show S_1 nuclease protection assays using 50000 cpm of ^{32}P-RNA (labeled by guanylyltransferase and alpha ^{32}P-GTP) and 100 fmoles of M13 mp18 containing the 1308 nuc BamHI-PstI fragment or 200 fmoles of the 1308 bp double stranded fragment respectively. The resulting assays show a major ^{32}P-RNA species that migrates at approximately 590 nucleotides. The assay shows that the single stranded DNA probe provides greater sensitivity in this assay. Lane M shows molecular weight markers.

could be some 'capped'-RNA species that were protected by residual ctDNA.

Correct initiation of transcription by E. coli

Since the promoter sequences of psbA and those of *E. coli* are similar, we investigated the possibility of proper transcription initiation of the psbA gene in *E. coli.* In order to test this hypothesis we contructed the plasmids pPPBX-10218 and pPPBX-10219. The plasmids have the 709 bp XbaI-PstI restriction fragment of pea ctDNA (that contains the entire putative promoter region, 5'-untranslated region, and 471 nuc of the psbA open reading frame) in opposite orientations with respect to the lacZ' gene of pUC18 and pUC19, respectively (see Fig. 6).

Total nucleic acid was isolated from *E. coli* that had been transformed with either pPPBX-10218 or pPPBX-10219. When *E. coli* total nucleic acid and ctRNA were hybridized to a 5'-end labeled XbaI-DdeI fragment (DdeI labeled) and treated with S_1 nuclease, fragments of similar size were protected from S_1 nuclease (Fig. 7). This suggests that *E. coli* RNA polymerase may be initiating transcription from the psbA promoter at a site similar to that used by the chloroplast RNA polymerase *in vivo.*

Transcription of the psbA gene in a homologous pea in vitro *transcription system*

The homologous *in vitro* transcription system used consisted of the plasmids pPPBX-10218 and pPPBX-10219 as DNA templates and chloroplast extracts which contain RNA polymerase. The RNA produced in these experiments along with the psbA RNA synthesized in *E. coli* were assayed by S_1 nuclease protection assays using the XbaI-HindIII probe shown in Fig. 6. This probe has 11 nucleotides of vector sequence which allowed us to distinguish between *in vivo* ctRNA, *in vitro* ctRNA and *E. coli* produced RNA. Figure 8 shows the results of the S_1 protection experiments where RNA synthesized *in vitro* from the psbA templates protected 565 nucleotides of the S_1 probe (lanes 2 and 4). The size of the *in vitro* RNA is similar to RNAs produced in *E. coli* which were transformed with the psbA-templates (lanes 3 and 5). In Fig. 7, it was shown that the psbA transcripts synthesized in *E. coli* protected psbA specific S_1 nuclease probes of similar size as in vivo ctRNA. Therefore the RNA transcripts produced in the homologous *in vitro* extracts were initiated from a similar site on the psbA template. The S_1 nuclease assays of the *in vitro* transcription reactions also show full-length and an

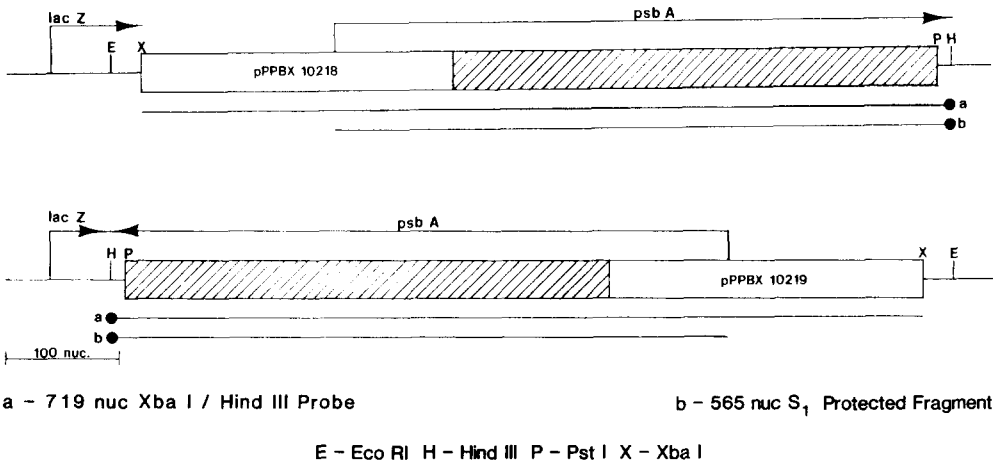


Fig. 6. The truncated psbA constructs used in analysis of transcription initiation. The top portion of this figure shows the plasmid pPPBX-10218. In this plasmid (pPPBX-10218) psbA transcription is in the same direction as transcription of the lacZ' gene of pUC18. In plasmid pPPBX-10219, psbA transcription is opposite that of the lacZ' gene of pUC19. The arrows show the direction and site of transcription initiation. The boxed regions are chloroplast derived sequences, with the cross-hatched areas being the psbA open-reading frame. The single lines are the vector sequences. 'a' is the 719 nuc XbaI-HindIII (5'-end labeled) prove and 'b' is the 565 nuc S_1 protected fragment (see Figs. 7 and 8).

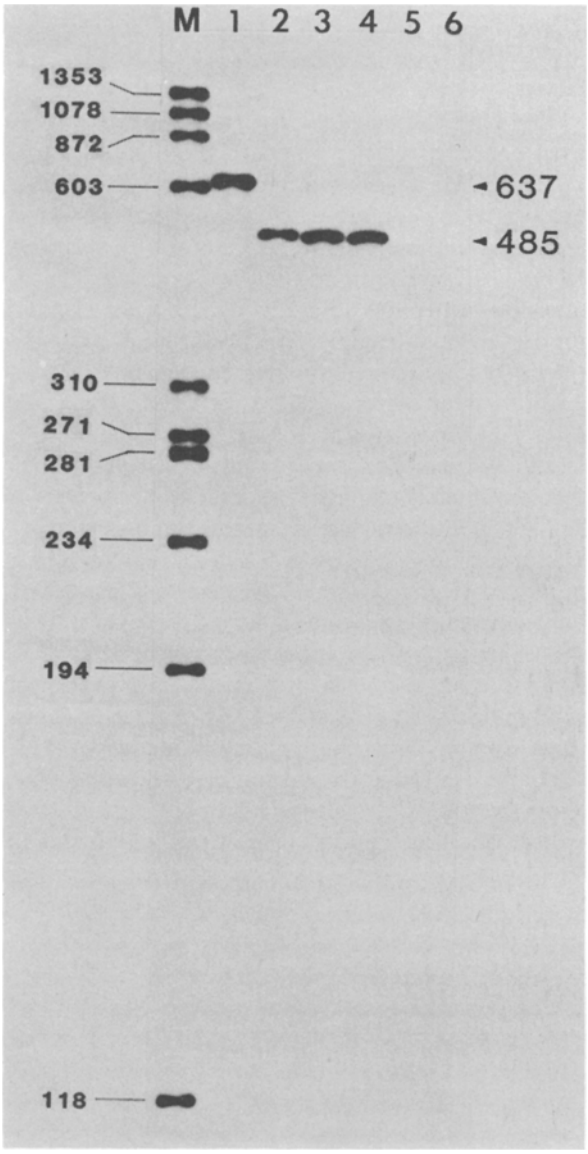


Fig. 7. S_1 nuclease assays of RNA synthesized in *E. coli* transformed with psbA. Lane 1 shows the 637 nuc XbaI-DdeI (5′-end labeled at the DdeI site) probe. Lane 2 shows an S_1 nuclease assay with the XbaI-DdeI probe and 0.021 μg of total nucleic acid from chloroplast. This lane shows the expected 485 nuc protected fragment. Lanes 3 and 4 contain 5 μg of total nucleic acid extracted from *E. coli* which had been transformed with the plasmids pPPBX-10218 and pPPBX-10219, respectively. Lanes 3 and 4 also show the characteristic 485 nuc protected fragment which suggests that *E. coli* RNA polymerase recognizes and initiates transcription near the psbA promoter. Lanes 5 and 6 contain the S_1 assays using 5 μg total nucleic acid extracted from *E. coli* which had been transformed with the plasmids pUC18 and pUC19, respectively. No bands are seen in lanes 5 and 6 indicating the vectors alone cannot generate psbA transcripts. The ∅X174-HaeIII (5′-end labeled) molecular weight markers are in lane M.

intermediate length S_1 protected band (Fig. 8, lanes 2 and 4). The full-length S_1 fragment was probably due to probe:probe hybridization caused by the addition of excess probe to the hybridization mixture. The intermediate fragment could be an S_1 artifact caused by AT rich regions of the probe:probe hybridization. The full-length S_1 protected fragment for the *E. coli*/pPPBX-10218 RNA (lane 3) could be due to transcription starting at the lacZ′ promoter and reading through the psbA sequences.

Control reactions with no chloroplast extract were used to assay for any residual (plasmid transformed) *E. coli* RNA or contaminating template after DNase I treatment. No bands were detected in the minus extract reaction (Fig. 8; lane 7). A control reaction in which the extract was present but template was not present, assayed for the discrimination of the *in vitro* probe for *in vitro* transcribed RNA (ie. the endogenous ctRNA would not protect the 5′-end label of the *in vitro* probe). This template control reaction also showed no bands (Fig. 8, lane 6). The pUC18 and pUC19 vectors used to construct the *in vitro* templates showed no S_1 nuclease protected bands when used in the *in vitro* system nor when RNA was isolated from pUC18 or pUC19 transformed *E. coli.*

Discussion

In this report we describe the characterization of *P. sativum* chloroplast psbA transcripts produced *in vivo, in vitro* and in *E. coli.* The 5′-end of psbA was previously localized to a 5.0 kb PstI DNA region (33) and the psbA open reading frame sequenced (31). We extended the previous sequence data upstream from the psbA open reading frame and located a putative 3′-exon of a gene coding for tRNA-Lys and an unidentified reading frame (Figs. 1, 2). The unidentified reading frame is the 3′-end of an open reading frame originating from a transcript whose 5′-end maps approximately 2.9 kbp upstream of the BamHI site in Fig. 1 (Boyer & Mullet, manuscript in preparation). The 89 amino acids of the unidentified reading frame reported here are highly charged and basic suggesting that the protein could be a nucleic acid binding protein. A homologous (70%) open reading frame and gene coding for a putative 3′-exon of tRNA-Lys

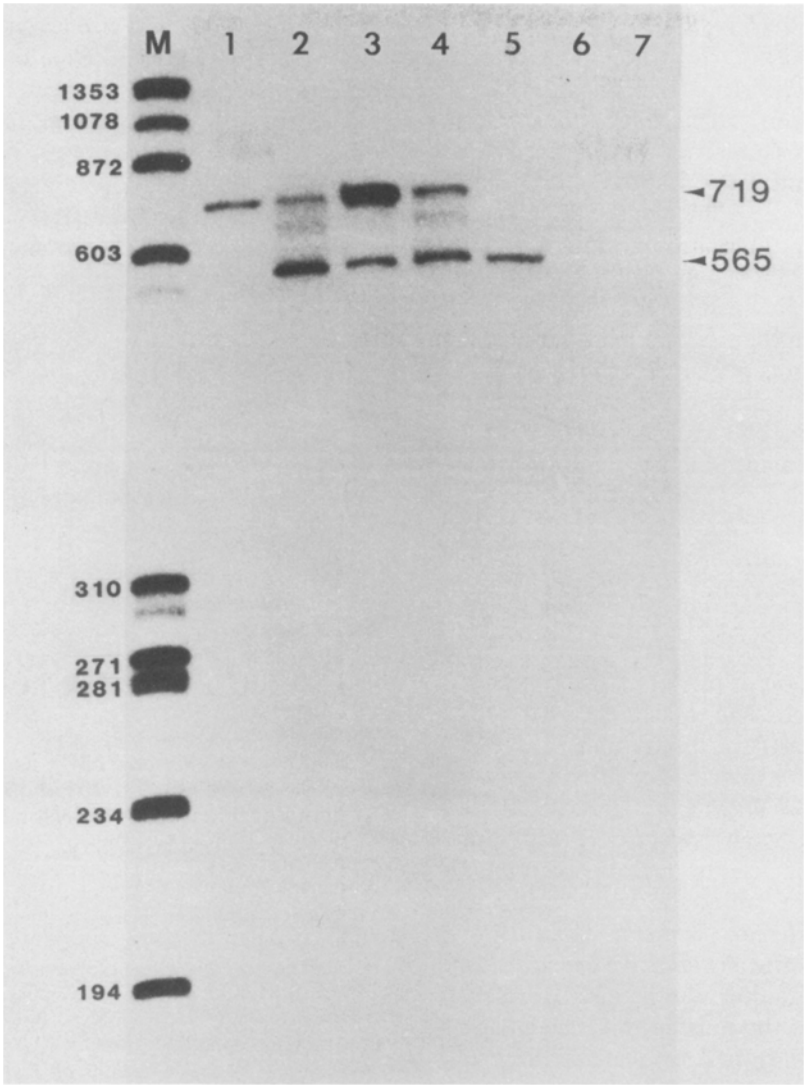


Fig. 8. S_1 nuclease assays of psbA RNA synthesized in a homologous chloroplast *in vitro* transcription system. Lane 1 shows the 719 nuc XbaI-HindIII (5′-end labeled at the HindIII site) S_1 nuclease protection probe. Lanes 2 and 4 are S_1 nuclease protection experiments of the RNA synthesized *in vitro* from pPPBX-10218 and pPPBX-10219 templates respectively. These lanes show a 565 nuc protected band which corresponds to initiation at the position of the *in vivo* psbA RNA 5′-end. Lanes 2 and 4 also show a full length protected band and a faint intermediate band. Lanes 3 and 5 are S_1 nuclease assays of RNA derived from *E. coli* transformed with pPPBX-10218 and pPPBX-10219 respectively. These lanes also show a 565 nuc protected band and lane 3 shows a full length protected fragment. Lane 6 is an S_1 nuclease protection assay of a control reaction where no template DNA was added to the *in vitro* system. Lane 7 is an S_1 nuclease protection assay were no chloroplast extract was added. The template DNA (pPPBX-10218) was assayed for any residual RNA contamination in this experiment and showed no protected bands. An analogous experiment was done with pPPBX-10219 and the results were the same as lane 7 (results not shown). Lane M is ∅X174-HaeIII molecular weight markers.

has been reported in *N. tabacum* (39). In *N. tabacum* evidence for splicing the tRNA-Lys was presented (39). A search for 3′-termini between the pea unidentified reading frame and psbA revealed transcript termini at the 3′-side of the putative 3′-exon of tRNA-Lys (Fig. 3a). This indicates that the tRNA is not processed to a large extent or that processed RNA is unstable. One other example of a protein coding region which resides in the intron of a tRNA has been published (8). In this case the

tRNA-Gly was shown to be processed (8).

Examination of psbA *in vivo* transcript 5'-ends revealed the presence of 5'-termini 92, 93 and 68 nucleotides upstream from the psbA open reading frame. The shorter transcript was present in low abundance. Transcripts with termini 92 – 93 nucleotides upstream from the psbA open reading frame were identified as primary transcripts. This was based on capping experiments using guanylyltransferase (Fig. 5), the presence of sequences 10 and 35 bp upstream from the termini that are homologous with prokaryotic promoter sequences (43), and production of transcripts similar to those found *in vivo, in vitro* and in *E. coli* (Figs 7, 8). It should be noted that production of psbA transcripts in *E. coli* which are similar to *in vivo* transcripts reported here is in contrast to results reported by Link (19). The difference may be due to our use of an *in vivo E. coli* assay in contrast to the *in vitro* assay used by Link which requires linear templates.

In vitro transcription of psbA in a pea transcription extract indicated that transcription of psbA was accurately initiated in this system. Link (19) has also reported an *in vitro* transcription system in mustard which can use linear templates to accurately synthesize 'run-off' transcripts of the psbA gene (19). Transcription of deletion mutants of psbA *in vitro* were used to determine the location and the requirement for different promoter elements for psbA (19). An element located 35 bp upstream from the site of initiation was required for transcription of psbA in the mustard system. Our sequence analysis of the pea psbA showed the presence of a similar sequence in pea psbA except that our sequence is GTGACA instead of TTGACA as found for mustard psbA. It should also be noted that our -35 sequence is a palindrome with sequence GGGTGACACCC. Palindromic sequences in *E. coli* are often found in regulatory regions and therefore this sequence may be a regulatory protein binding site in the pea chloroplast.

DNA sequences between the -35 region of psbA and the upstream URF 3'-end could be involved in the regulation of psbA transcription. Sequence comparison between *P. sativum* and *N. tabacum* (39) in this region show a lack of homology between the sequence 40 bp upstream from psbA RNA 5'-end and the sequence demarked by the 3'-terminus of the tRNA-Lys sequence (Fig. 2). The lack of homology and size variation (pea is 199 bp and tobacco is 88 bp) of this region suggests that regulatory sequences, if present, have not been conserved between these species. The divergent intergenic spacer region, however, is very AT rich in both species and may be a site of entry for RNA polymerase. Further study of regions between other genes in the chloroplast genome should aid in the determination of the function of these intergenic regions.

Acknowledgement

Supported in part by grant DCB-8316431 to JEM from the NSF Developmental Biology Program.

References

1. Baldwin TO: *E. coli* TB1-host for pUC plasmids. Focus 6/4:7, 1984.
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. Bedbrook JR, Link G, Coen DM, Bogorad L, Rich A: Maize plastid gene expressed during photoregulated development. Proc Natl Acad Sci USA 75:3060 – 3064, 1978.
4. Blair G, Ellis R: Protein synthesis in chloroplasts: I. Light-driven synthesis of the large subunit of fraction I protein by isolated pea chloroplasts. Biochim Biophys Acta 319:223 – 234, 1973.
5. Bujard H: The interactions of *E. coli* RNA polymerase with promoters. Trends Biochem Sci 5:274 – 278, 1980.
6. Chu NM, Shapiro DR, Oishi KK, Tewari KK: Distribution of transfer RNA genes in the *Pisum sativum* chloroplast DNA. Plant Mol Biol 4:65 – 79, 1985.
7. Crossland LD, Rodermel SR, Bogorad L: Single gene for the large subunit of ribulosebisphosphate carboxylase in maize yields two differently regulated mRNAs. Proc Natl Acad Sci USA 81:4060 – 4064, 1984.
8. Deno H, Sugiura M: Chloroplast tRNA-Gly gene contains a long intron in a D stem: Nucleotide sequences of tobacco chloroplast genes for tRNA-Gly (UCC) and tRNA-Arg (UCU). Proc Natl Acad Sci USA 81:405 – 408, 1984.
9. Frankfort HM, Robertson HD: In: Linn SM, Roberts RJ (eds) Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982, pp 359 – 366.
10. Fromm H, Devic M, Fluhr R, Edelmann M: Control of psbA gene expression in mature Spirodela chloroplasts light regulation of 32 kd protein synthesis is independent of transcript level. EMBO J 4:291 – 295, 1985.

11. Gruissem W, Greenburg BN, Zurawski G, Prescott DM, Hallick RB: Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. Cell 35:815 – 828, 1983.
12. Gruissem W, Zurawski G: Identification and mutational analysis of the promoter for a spinach transfer RNA gene. EMBO J 4:1637 – 1644, 1985.
13. 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.
14. Hallick RB, Hollingsworth MJ, Nickdoff JA: Transfer RNA genes of *Euglena gracilis* chloroplast DNA. Plant Mol Biol 3:169 – 175, 1984.
15. Herrmann RG, Westhoff P, Alt J, Tittgen J, Nelson N: Thylakoid membrane proteins and their genes. In: Vloten-Doting LV, Grant GSP, Hall TC (eds) Molecular Form and Function of the Plant Genome, Plenum Publishing Corp. 1985, pp 233 – 256.
16. Hirschberg J, McIntosh L: Molecular basis of herbicide resistance in *Amaranthus hybridus*: Science 222:1346 – 1349, 1983.
17. Hirschberg J, Bleeker A, Kyle DJ, McIntosh L, Arntzen CJ: The molecular basis of triazine herbicide resistance in higher plant chloroplast. Z Naturforsch 39c:412 – 420, 1984.
18. Link G: Phytochrome control of plastid messenger RNA in mustard (*Sinapsis alba. L.*). Planta 154:81 – 86, 1982.
19. Link G: DNA sequence requirement for the accurate transcription of protein-coding plastid gene in a plastid *in vitro* system from mustard (*Sinapsis alba. L.*). EMBO J 3:1697 – 1704, 1984.
20. Link G: Hybridization study of developmental plastid gene expression in mustard (*Sinapsis alba. L.*) with cloned probes for most plastid DNA regions. Plant Mol Biol 3:243 – 248, 1984.
21. Link G, Langride U: Structure of the chloroplast gene for the precursor of the M_r 32000 photosystem II protein from mustard (*Sinapsis alba L.*) Nuc Acids Res, 12:945 – 958, 1984.
22. Lizardi PM: Schleicher and Schuell publication #364, 1983.
23. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982.
24. Mattoo AK, Pick U, Hoffman-Folk H, Edelman M: The rapidly metabolized 32000 dalton polypeptide of the chloroplast is the 'proteinaceous shield' regulating photosystem II electron transport and mediating diuron herbicide sensitivity. Proc Natl Acad Sci USA 78:1572 – 1576, 1981.
25. 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, 1982, pp 499 – 560.
26. Messing J: New M13 vectors for cloning. In: Colowich SP, Kaplan NO (eds) Methods in Enzymology, Vol. 101C. Academic Press, New York, 1980, pp 20 – 77.
27. 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.
28. Mullet JE, Orozco EM, Chua N-H: Multiple transcripts for higher plant rbcL and atpB genes and localization of the transcription initiation site of the rbcL gene. Plant Mol Biol 4:39 – 54, 1985.
29. Nelson T, Harpster MH, Mayfield SP, Taylor WC: Light-regulated gene expression during maize leaf development. J Cell Biol 98:558:564, 1984.
30. Norander J, Kempe T, Messing J: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101 – 106, 1983.
31. Oishi KK, Shapiro DR, Tewari KK: Sequence organization of a pea chloroplast DNA gene coding for a 34500-dalton protein. Mol Cell Biol 4:2556 – 2563, 1984.
32. Orozco EM, Mullet JE, Chua N-H: An *in vitro* system for accurate transcription initiation of chloroplast protein genes. Nuc Acids Res 13:1283 – 1302, 1985.
33. Palmer JD, Thompson WF: Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537 – 550, 1982.
34. Pfister K, Steinback KE, Gardner G, Arntzen CH: Photoaffinity labeling of an herbicide receptor protein in chloroplast membranes. Proc Natl Acad Sci USA 78:981 – 985, 1981.
35. Rodermel SR, Bogorad L: Maize plastid photogenes: mapping and photoregulation of transcript levels during light induced development. J Cell Biol 100:463 – 476, 1985.
36. Spielman A, Stutz E: Nucleotide sequence of soybean chloroplast DNA regions which contain the psbA and trnH genes and cover the ends of the large single copy region and one end of the inverted repeats. Nuc Acids Res 11:7157 – 7167, 1983.
37. Steinback K, McIntosh L, Bogorad L, Arntzen C: Identification of the triazine receptor protein as a chloroplast gene product. Proc Natl Acad Sci USA 78:7463 – 7467, 1981.
38. Sugita M, Sugiura M: Nucleotide sequence of the gene for the 32000 dalton thylakoid membrane protein from *N. tabacum*. Mol Gen Genet 195:308 – 313, 1984.
39. Sugita M, Shinozaki K, Sugiura M: Tobacco chloroplast $tRNA^{Lys}$ (UUU) gene contains a 2.5-kilobase pair intron: An open reading frame and a conserved boundary sequence in the intron. Proc Natl Acad Sci USA 82:3557 – 3561, 1985.
40. Thompson WF, Everett M, Polans NO, Jorgensen RA: Phytochrome control of RNA levels in developing pea and mung bean leaves. Planta 158:487 – 500, 1983.
41. Tullis RH, Rubin H: Calcium protects DNase I from Proteinase K: A new method for removal of contaminating RNase from DNase I. Anal Biochem 107:260 – 264, 1980.
42. Whitfeld PR, Bottomley W: Organization and structure of chloroplast genes. Ann Rev Plant Physiol 34:279 – 310, 1983.
43. Youderian P, Bouvier S, Susskind MM: Sequence determinants of promoter activity. Cell 30:843 – 853, 1982.
44. Zurawski G, Bohnert HJ, Whitfeld PR, Bottomley W: Nucleotide sequence of the Mr 32000 thylakoid membrane protein from *Spinacia oleracea* and *Nicotiana debneyi* predicts a totally conserved primary translation product of Mr 38950. Proc Natl Acad Sci USA 79:7699 – 7703, 1982.

Received 27 June 1985; in revised form 11 December 1985; accepted 16 December 1985.