Volume 13 Number 4 1985 Nucleic Acids Research

# An *in vitro* system for accurate transcription initiation of chloroplast protein genes

Emil M.Orozco, Jr.*, John E.Mullet+ and Nam-Hai Chua

Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA

Received 10 August 1984; Revised and Accepted 28 January 1985

## ABSTRACT

We have developed an homologous in vitro system from spinach chloroplasts that correctly initiates transcription of the plastid genes for the large subunit of ribulose-1,5-bisphosphate carboxylase (rbcL) and the β subunit of the plastid ATPase (atpB). The transcriptionally active extracts from spinach chloroplasts require circular DNA templates for specific initiation. The RNA polymerase activity is insensitive to rifampicin. The extent of transcription in vitro is a function of the extract:template ratio. The efficiency of the rbcL transcription in vitro is more than one transcript per one hundred templates per hour.

## INTRODUCTION

In higher plants the circular plastid genome has a size of 120 to 180 kbp and contains the genes for plastid ribosomal and transfer RNAs as well as various proteins involved in photosynthesis (1,2). Although many of these plastid genes have been isolated and their DNA sequences determined little is known concerning the nature of their transcriptional regulation. Chloroplast enzymes are typically oligomers consisting of both plastid and nuclear-encoded subunits. In spinach, the plastid DNA is present in 4000 to 4500 copies per leaf cell depending on the particular developmental stage of the plastid (3). The nuclear and plastid genetic systems must interact in some manner to express coordinately the low copy number nuclear genes (4) with the much higher copy number plastid genes. A characterization of the transcriptional properties of plastid protein-coding genes should provide a first step toward understanding the nature of this interaction.

Various investigators (1, 2) have observed that 5' to many plastid genes are sequences similar to the "-10" and "-35" regions of the E. coli consensus promoter sequence (5, 6). The physiological significance of these sequences in chloroplast gene transcription have not yet been determined. In order to understand the transcriptional regulation of plastid protein-coding genes we are interested in defining which DNA sequence elements are required

for transcription regulation and if polypeptide factors other than RNA polymerase interact with these gene regulatory elements. To address these issues it was necessary to develop a cell-free system which would accurately initiate transcription of exogenously added template. In this report we describe the development and partial characterization of an homologous *in vitro* transcription system from spinach chloroplasts. As a paradigm, we have examined the *in vitro* transcription of the plastid genes for the large subunit of the $CO_2$-fixing enzyme ribulose-1,5-bisphosphate carboxylase (*rbcL*) and the β subunit of the plastid ATPase (*atpB*). Under the conditions described in this report the spinach chloroplast transcription extracts will correctly initiate transcription of both of these genes. The availability of this chloroplast *in vitro* transcription system should allow for the characterization of the promoter structure of these and other plastid protein-coding genes as well as the isolation of putative factors involved in plastid gene regulation.

## MATERIALS AND METHODS

### Reagents and Enzymes

Rifampicin and *E. coli* tRNA type XXI were from Sigma Chemical Co. Trace proteins in tRNA samples were removed by phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol) extractions. ØX174-*Hae*III DNA fragments (New England Biolabs Inc.) were 5' end labeled with $\gamma$-$^{32}$P-ATP and T4 polynucleotide kinase and used as molecular size standards. Nucleic acid-grade formamide (BRL, Inc.) was deionized for 15 minutes by treatment with AG501-X8 (D) analytical grade mixed bed resin (Bio-Rad Laboratories) and stored at -20°C. Preswollen DE-52 resin was from Whatman Chemical Separation Ltd. Calf intestine alkaline phosphatase and proteinase K were from Boehringer-Mannheim. S1 nuclease and *E. coli* RNA polymerase were from BRL, Inc. DNaseI was from Worthington Diagnostic Systems, Inc.

### Plant Growth Conditions

*Spinacea oleracea* (hybrid No. 424) was purchased from the Ferry Morse Seed Co. and grown in vermiculite for seven days in a Conviron environmental growth chamber set for 12 hour days (22°C) and 12 hour nights (18°C).

### Chloroplast RNA Isolation

Intact chloroplasts were isolated after centrifugation through percoll gradients as described by Bartlett *et al.* (7) except that 40-85% step gradients were used rather than 10-80% linear gradients. Purified plastids were lysed with an equal volume of 6M urea, 0.36M 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). Chloroplast RNAs were collected from the aqueous phase by precipitation twice from ethanol, resuspended in distilled deionized water and stored at -80°C.

Plasmid DNA Isolation

Plasmid DNAs were isolated by the alkaline-SDS method (8) and further purified by CsCl-ethidium bromide gradients (9). The pSoc801 and pSoc802 plasmid DNAs were typically greater than 50% supercoiled and were present to a variable but significant percentage as dimer and higher order oligomers. Before use as templates for *in vitro* transcription reactions, the plasmid DNAs were additionally treated with 200 μg/ml of proteinase K at 37°C for 2 hours, and the enzyme subsequently removed by phenol:chloroform:isoamyl alcohol extractions.

Isolation of Radioactively Labeled Single-Stranded DNA

Plasmid DNA was digested with the appropriate restriction enzymes and then electrophoresed through agarose gels. The appropriate section of the gel was excised, the DNA fragment isolated by electroelution into dialysis tubing (10) and then further purified by DE-52 column chromatography at room temperature. After treatment with calf intestine alkaline phosphatase (10) the DNA fragment was 5' end labeled with $\gamma$-$^{32}$P-ATP and T4 polynucleotide kinase (11). The $^{32}$P-DNA was then denatured, the two strands separated on a 5% (30:0.6; acrylamide:bisacrylamide) polyacrylamide gel and the single-stranded DNAs isolated as described by Maxam and Gilbert (11).

Preparation of Transcriptionally Active "High-Salt Extracts" from Spinach Chloroplasts

All manipulations were performed at 4°C. Intact chloroplasts were purified on percoll step gradients as described above. Chloroplasts were pelleted for five minutes at 6500 x g and resuspended in 330 mM sorbitol, 50 mM Hepes-KOH, pH 8, 5 mM sodium ascorbate, 2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, to approximately 2 x $10^9$ plastids/ml. Five volumes of "lysis buffer" (20 mM Hepes-KOH, pH 8, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine, 5 mM $\varepsilon$-amino-n-caproic acid) were added and the solution mixed by gentle inversions. The lysed chloroplast membranes were pelleted at 6500 x g for five minutes. The supernatant was discarded and the pellet was resuspended in "extraction buffer" (lysis buffer containing 15% glycerol and 1M NaCl) to an equivalent of 2 x $10^9$ plastids/ml. This mixture was gently stirred for 30 minutes and then centrifuged for 30 minutes at 80,000 x g. The supernatant was removed and 0.32 g of ammo-

nium sulfate (BRL ultra-pure reagent grade) were added per ml of extract. The mixture was gently mixed by occasional inversions for 60 minutes. The precipitated protein was then collected by centrifugation for 15 minutes at 60,000 x g. The supernatant was discarded and the pellet was resuspended in a minimal volume of "DE-52 buffer" (50 mM Tricine-KOH, pH 8, 50 mM KCl, 0.5 mM EDTA, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine, 5 mM ε-amino-n-caproic acid, 5% glycerol). The mixture was dialyzed overnight versus two changes of DE-52 buffer. This high-salt extract was then subjected either to DEAE column chromatography or was divided into aliquots, frozen in liquid nitrogen, and stored at -80°C. The high-salt extract will specifically initiate transcription after storage at -80°C for 14 months.

Preparation of the DEAE Enzyme Fraction

All manipulations were performed at 4°C. After dialysis, 0.5 ml of the high-salt extract (obtained from plant tissue containing 2.6 mg of chlorophyll) was applied to a 0.4 ml DEAE column. The column was washed successively with 0.5 ml fractions of DE-52 buffer containing the following concentrations of KCl: (1) 200 mM, (2) 300 mM, (3) 400 mM, and (4) 500 mM. Each fraction was dialyzed and stored as described above. The fraction eluting with 300 mM KCl is the DEAE enzyme fraction used in Figure 2. A more complete separation of enzyme from nucleic acids was obtained in a separate experiment by elution from DEAE with a linear KCl gradient from 50-550 mM. In this preparation plant tissue containing 42 mg of chlorophyll was used. This second DEAE enzyme fraction was used for the experiment shown in Figure 3. The DEAE enzyme fraction will specifically initiate transcription after storage at -80°C for three months.

RNase-Free DNase

RNase-free DNase was prepared by incubation of DNaseI with proteinase K in the presence of calcium ions. Calcium ions cause the DNaseI to become resistant to digestion by proteinase K (12). DNaseI was resuspended in 20 mM Tris-HCl, pH 7.4, 10 mM $CaCl_2$ to a final concentration of 1.11 mg/ml and then incubated at 37°C for 5 minutes. To this solution was added one-ninth volume of a 10 mg/ml solution of proteinase K and the enzymes were incubated at 37°C for 30 minutes. The DNaseI-proteinase K mixture was then immediately used to digest DNA as described below.

Analysis of In Vitro Transcripts

Each in vitro transcription reaction (20-45 µl) contained 12 mM Hepes-KOH, pH 8, 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 500 µM GTP, 500 µM CTP, 50 µM ATP, and 50 µM UTP as well as the following components: one-fifth of the

volume was proteinase K-treated DNA in 10 mM Tris-HCl, pH 8, 0.1 mM EDTA and two-fifths of the reaction consisted of DE-52 buffer with variable amounts of either high-salt extract or DEAE enzyme fraction (each of which had been previously dialyzed against DE-52 buffer). The DE-52 buffer and extracts contributed an additional 20 mM of KCl to the final reaction volume. The *in vitro* transcription reaction was incubated at 30°C for either 30 or 60 minutes. DNA was then removed from the *in vitro* transcription reaction by adding the following: 15 µl of 200 mM Tris-HCl, pH 7.4, 100 mM $MgCl_2$, 35 mM $CaCl_2$; 10 µl of DNaseI-proteinase K mixture; and $H_2O$ to a final volume of 150 µl. This reaction was incubated at 37°C for 30 minutes and then stopped by the addition of 150 µl of "RNA extraction buffer" (6M urea, 0.36M NaCl, 20 mM EDTA, 10 mM Tris-HCl, pH 8, 1% sodium dodecylsulfate). *E. coli* tRNA (15 µg/reaction) was added as carrier and the *in vitro* RNA extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1; vol:vol:vol), ethanol precipitated twice, the pellet rinsed once with cold (-20°C) 80% ethanol, dried and then resuspended in distilled deionized water and stored at -80°C. The RNA was then assayed by S1 nuclease protection as described previously (13) based on the method of Weaver and Weissmann (14).

Quantitation of Autoradiogram Data

Bands on autoradiograms were excised and the silver grains were eluted and quantitated as described by Suissa (15).

## RESULTS

### Description of the Chloroplast DNA Template

We have chosen to investigate *in vitro* transcription of chloroplast protein genes using the spinach *rbcL* and *atpB* genes as a model system. These two genes are separated by 152 bp and are transcribed divergently (16, 17). Moreover, their complete nucleotide sequences have been elucidated (16, 17, 18) and the 5' ends of their *in vivo* transcripts have been mapped (13, 16). Each gene has several *in vivo* transcripts (Figure 1). The larger *rbcL* transcript has been shown by *in vitro* "capping" analysis to be a primary transcription product (13). The 1041 bp *MspI-PstI* DNA fragment shown in Figure 1 was cloned into the vectors pUC8 and pUC9 (19) to obtain pSoc801 and pSoc802 (13). These constructions differ only in the orientation of the insert DNA relative to the vector DNA sequence. The pSoc801 and pSoc802 DNAs were used in this study as templates for *in vitro* transcription.

### The Homologous Spinach In Vitro Transcription System Will Correctly Transcribe Exogenous DNA Templates

Our *in vitro* transcription system consists of two major components: (1)

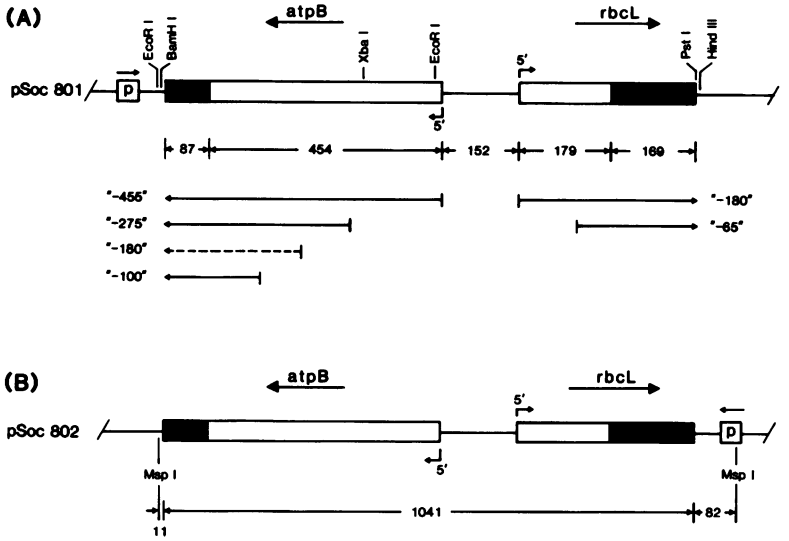


Figure 1. A schematic diagram of the spinach chloroplast DNA fragment that contains 5' regions of the rbcL and atpB genes. The 1041 bp MspI-PstI DNA fragment from spinach chloroplast DNA was cloned into the AccI and PstI sites of pUC8 and pUC9 to obtain pSoc801 and pSoc802, respectively (13). A. Orientation of the MspI-PstI chloroplast DNA in pSoc801. The lac promoter ("P") and direction of lacZ transcription are indicated. The shaded boxes indicate protein coding regions and the open boxes indicate transcribed but nontranslated regions. The distances (bp) of each of these regions and the intergenic spacer are also given. The different transcripts for each gene have been identified previously (13) and are represented here by arrows. Each transcript is identified by the position of its 5' end relative to its respective open reading frame. The transcript represented by a dotted line (atpB "-180") is a minor species in vivo (13). B. Orientation of the MspI-PstI spinach chloroplast DNA fragment in pSoc802 with respect to the lac promoter ("P"). The directions of transcription of the atpB, rbcL, and lacZ genes are indicated. Also indicated is the 1134 bp MspI DNA fragment from pSoc802 that was used as a probe for S1 nuclease protection analysis. The size (bp) of vector DNA between the MspI sites and the spinach chloroplast DNA is indicated. This probe is specific for S1 nuclease protection by in vitro synthesized RNAs (see text).

pSoc801 and pSoc802 DNA as templates and (2) chloroplast extracts that have transcriptional activity. In preliminary experiments, we used transcriptionally active high-salt extracts and assayed the in vitro transcripts by S1 nuclease protection. However, these extracts yielded significant S1 nuclease protection in the absence of added template. This background signal could have been due to contaminating chloroplast mRNA in the extracts and/or in vitro transcription of contaminating chloroplast DNA. To eliminate this background problem, we removed most of the contaminating nucleic

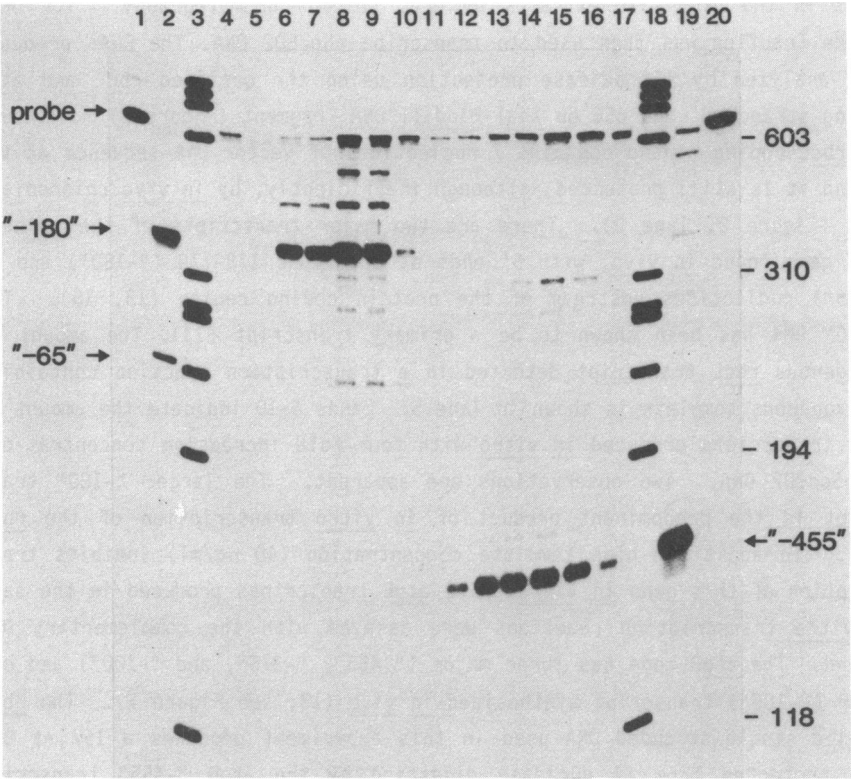


Figure 2. Correct transcription initiation in vitro of the rbcL and atpB genes. The in vivo and in vitro synthesized RNAs were assayed by S1 nuclease protection analysis using either the purified rbcL-coding strand (lanes 2 and 4-10) or the purified atpB-coding strand (lanes 11-17 and 19) of the 656 bp XbaI-HindIII DNA of pSoc801 (7500 cpm/assay). Lane 1: 100 cpm of the rbcL-coding strand. Lane 2: S1 nuclease protection of the rbcL-coding strand by 0.2 µg of chloroplast RNA. Lanes 3 and 18: The ØX174-HaeIII DNA fragments used as DNA size standards. Lane 19: S1 nuclease protection of the atpB-coding strand by 2.0 µg of chloroplast RNA. Lane 20: 100 cpm of the atpB-coding strand. For each in vitro transcription reaction (40 µl), 45% was analyzed for rbcL transcripts (lanes 4-10) and 45% was analyzed for atpB transcripts (lanes 11-17). Lanes 4 and 11: Analysis of a mock transcription reaction that contained 40 µg/ml of pSoc802 DNA and no DEAE enzyme fraction. The remaining transcription reactions each contained 4 µl of DEAE enzyme fraction and the indicated amount of pSoc802 DNA. Lanes 5 and 12: No DNA. Lanes 6 and 13: 0.16 µg/ml. Lanes 7 and 14: 0.62 µg/ml. Lanes 8 and 15: 2.5 µg/ml. Lanes 9 and 16: 10 µg/ml. Lanes 10 and 17: 40 µg/ml. The sizes (in nucleotides) of four of the ØX174-HaeIII DNAs are shown to the right of lane 20. The positions of the DNAs that are protected by rbcL ("-180" and "-65") and atpB ("-455") transcripts are also indicated.

acids in the high-salt extracts by DEAE column chromatography. The DEAE enzyme fraction was then used to transcribe pSoc802 DNA. The RNAs produced were analyzed by S1 nuclease protection using the purified rbcL and atpB coding strands of the 656 bp XbaI-HindIII DNA Fragment (Figure 2). Although the rbcL-coding strand contains 7 nucleotides of vector DNA sequence at the 5' end it is still protected, although inefficiently, by in vivo chloroplast RNAs (Figure 2, lane 2). There are two major transcripts of the spinach rbcL gene found in vivo, with 5' ends at positions 178-179 ("-180") and 64 ("-65") nucleotides upstream of the protein coding region (13, 16). The "-180" RNA has been shown to be a primary transcript (13). The amount of endogenous rbcL transcript detected in a transcription reaction containing no exogenous template is shown in lane 5. Lanes 6-10 indicate the amount of rbcL transcripts produced in vitro with four-fold increasing concentrations of pSoc802 DNA. Two observations are apparent. The larger "-180" transcript is the predominant product of in vitro transcription of the rbcL gene. In addition, high template concentration (40 µg/ml) inhibits transcription of this gene in vitro. The atpB transcripts produced in the same in vitro transcription reactions were assayed with the complementary DNA strand. The atpB gene has three major ("-455", "-275", and "-100") and one minor ("-180") transcript synthesized in vivo (13; see Figure 1). The XbaI labeled single-stranded DNA used in this experiment produces a 157 nt DNA when protected from S1 nuclease digestion by the atpB "-455" transcript (Figure 2, lane 19). The production of this transcript in vitro is shown in lanes 11-17. The amount of endogenous atpB transcript present in a reaction containing no exogenously added template is shown in lane 12. The atpB single-stranded DNA probe contains no vector DNA sequences, and consequently, this endogenous signal is greater than that seen in lane 5 for the rbcL gene. The amount of atpB transcription produced in vitro with increasing concentrations of template is shown in lanes 13-17. As with the rbcL gene, transcription of the atpB gene in vitro is inhibited by high template concentration (40 µg/ml) and the predominant in vitro transcript observed has its 5' end at the position of the largest in vivo transcript ("-455"). S1 protection analysis of this gene with longer DNA probes shows very little in vitro transcription initiation in the remainder of the long leader region before the atpB protein coding region (see below).

Transcription In Vitro is Independent of Vector DNA Sequences

To ensure that vector DNA sequences have no effect on gene transcription in vitro, the templates activities of pSoc801 and pSoc802 DNAs were

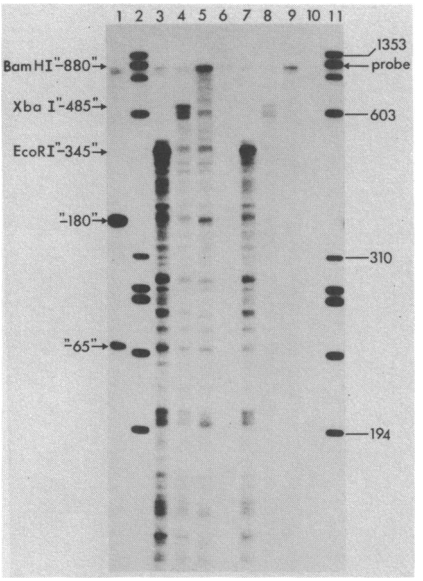


Figure 3. Transcription initiates non-specifically when linear DNAs are used as templates. Spinach chloroplast RNA and in vitro transcribed RNAs were assayed by S1 nuclease protection analyses using 15,000 cpm/assay of the single-stranded 1052 nt BamHI-HindIII DNA (5' end labeled at the HindIII site). Lane 1: S1 nuclease protection assay with 0.06 μg of spinach chloroplast RNA. Protection of DNA due to "-180" and "-65" rbcL mRNAs is shown to the left of lane 1. Lanes 2 and 11: The ØX174-HaeIII DNA fragments used as DNA size standards. The sizes (in nucleotides) of four of these standards are shown to the right of lane 11. The size of the BamHI-HindIII DNA used as S1 probe is also shown. Lanes 3-10: S1 nuclease protection of in vitro synthesized RNAs. Each transcription reaction contained 3 μl of the DEAE enzyme fraction in a final volume of 45 μl and 2.7 μg/ml of the following DNAs: EcoRI-digested pSoc801 (lane 3); XbaI-digested pSoc801 (lane 4); BamHI-digested pSoc801 (lane 5); HindIII-digested pSoc801 (lane 6); EcoRIdigested pSoc802 (lane 7); XbaI-digested pSoc802 (lane 8); BamHI-digested pSoc802 (lane 9); and HindIII-digested pSoc802 (lane 10). The signal expected from transcripts initiating at the end of each linear DNA is indicated to the left of lane 1 (see also Figure 1).

compared. These two templates differ only in the orientation of the chloroplast DNA fragment with respect to the pUC vector DNA (Figure 1). The same transcripts are made when either of these two DNAs are used as templates (data not shown) demonstrating that rbcL and atpB transcription in vitro initiates independently of the vector DNA sequences.

Transcription Initiation In Vitro is Nonspecific with Linear DNA Templates

The in vitro systems for RNA polymerase II genes often employ DNA templates linearized at a site a few hundred bp downstream from the transcription start site (20, 21). Such truncated templates were found to generate $^{32}$P-labeled "run-off" transcripts of a discrete size which could then be assayed directly by polyacrylamide gel electrophoresis. To see whether similar truncated templates could be used in our system we isolated various restriction fragments that contained the 5' ends of the rbcL and atpB genes. Transcription of these linear templates in the presence of $\alpha$-$^{32}$P-UTP produced a heterogeneous population of RNAs of total fragment length as well

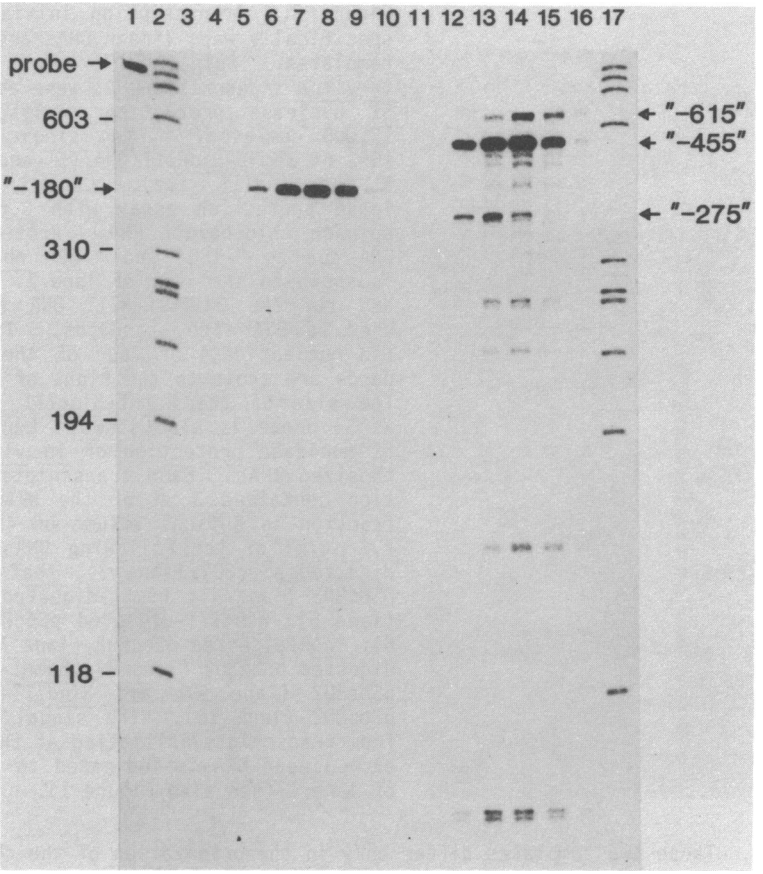


Figure 4. Transcription in vitro of the rbcL and atpB genes by the high-salt extract. The in vitro synthesized RNAs were assayed by S1 nuclease protection analysis using either the purified rbcL-coding strand (lanes 3-9) or the purified atpB-coding strand (lanes 10-16) of the 1134 bp MspI DNA from pSoc802 (15,000 cpm/assay). These single-stranded S1 probes are specific for the in vitro synthesized transcripts (see text). Lane 1: 400 cpm of the rbcL-coding strand of the 1134 bp MspI DNA. Lanes 2 and 17: The ØX174-HaeIII DNA fragments used as DNA size standards. For each in vitro transcription reaction (40 µl), 10% was analyzed for rbcL transcripts (lanes 3-9) and 30% was analyzed for atpB transcripts (lanes 10-16). Lanes 3 and 10: Mock transcription reaction containing no high-salt extract and 40 µg/ml of pSoc802 DNA. The remaining reactions each contained 4 µl of high-salt extract and the indicated amount of pSoc802 DNA. Lanes 4 and 11: No DNA. Lanes 5 and 12: 0.16 µg/ml. Lanes 6 and 13: 0.62 µg/ml. Lanes 7 and 14: 2.5 µg/ml. Lanes 8 and 15: 10 µg/ml. Lanes 9 and 16: 40 µg/ml. The positions of the DNAs that are protected by in vitro synthesized rbcL ("-180") and atpB ("-615", "-455" and "-275") transcripts are indicated. The sizes (in nucleotides) of four of the ØX174-HaeIII DNAs are shown to the left of lane 1.

as smaller sizes (data not shown). To examine the nature of this transcript heterogeneity we digested pSoc801 and pSoc802 DNAs with each of the restriction enzymes EcoRI, XbaI, BamHI, and HindIII. These enzymes restrict the plasmid DNAs at positions "-345", "-485", "-880", and "+170" relative to the rbcL translation start site, respectively. The EcoRI enzyme also restricts the plasmid DNAs in the multiple cloning site at position "-890" relative to the rbcL protein coding region (Figure 1). These linear DNAs were used as templates for in vitro transcription and the resulting RNAs were examined by S1 nuclease protection (Figure 3). No transcripts were detected for the templates restricted at the HindIII site. However, the $^{32}$P-DNA used for S1 nuclease protection was the 1052 nt BamHI-HindIII DNA, 5' end labeled at the HindIII site. Putative "run-off" transcripts from a HindIII restricted template may not have been long enough to adequately protect the $^{32}$P-HindIII end from S1 nuclease digestion. For the other linear DNA templates very little transcription initiation was observed at the "-180" position of the rbcL gene. Instead, transcripts were found to initiate at the site where each DNA template was linearized and, to a variable extent, at other nonspecific sites throughout the template.

In Vitro Transcription of the rbcL and atpB Genes by the High-Salt Extract.

The chloroplast DEAE enzyme fraction initiates transcription at the correct rbcL and atpB start sites. However, it would also be advantageous to study chloroplast gene transcription with the high-salt extract. For example, this cruder enzyme preparation may contain regulatory factors that do not copurify with the RNA polymerase after DEAE chromatography. For the experiments shown in Figures 4-6 we employed the chloroplast high-salt extract instead of the DEAE enzyme fraction. To discriminate between endogenous RNA and transcripts synthesized in vitro, we selected as probe for S1 nuclease protection the 1134 bp MspI fragment from pSoc802 (Figure 1). Since this DNA contains 82 and 11 nt of pUC9 DNA at the rbcL and atpB termini, respectively, the $^{32}$P-label will only be protected by transcripts of the recombinant DNA template and not by endogenous rbcL or atpB mRNA. In addition, this longer S1 probe should detect any transcription initiation events occurring in the 455 bp long atpB leader region. Transcription of increasing amounts of template DNA by the high-salt extract is depicted in Figure 4. As expected, no endogenous signal is detected for a mock transcription reaction containing high-salt extract and no exogenously added template DNA (lanes 4 and 11 for rbcL and atpB, respectively). In this experiment, as in the experiment shown in Figure 2, the predominant tran-

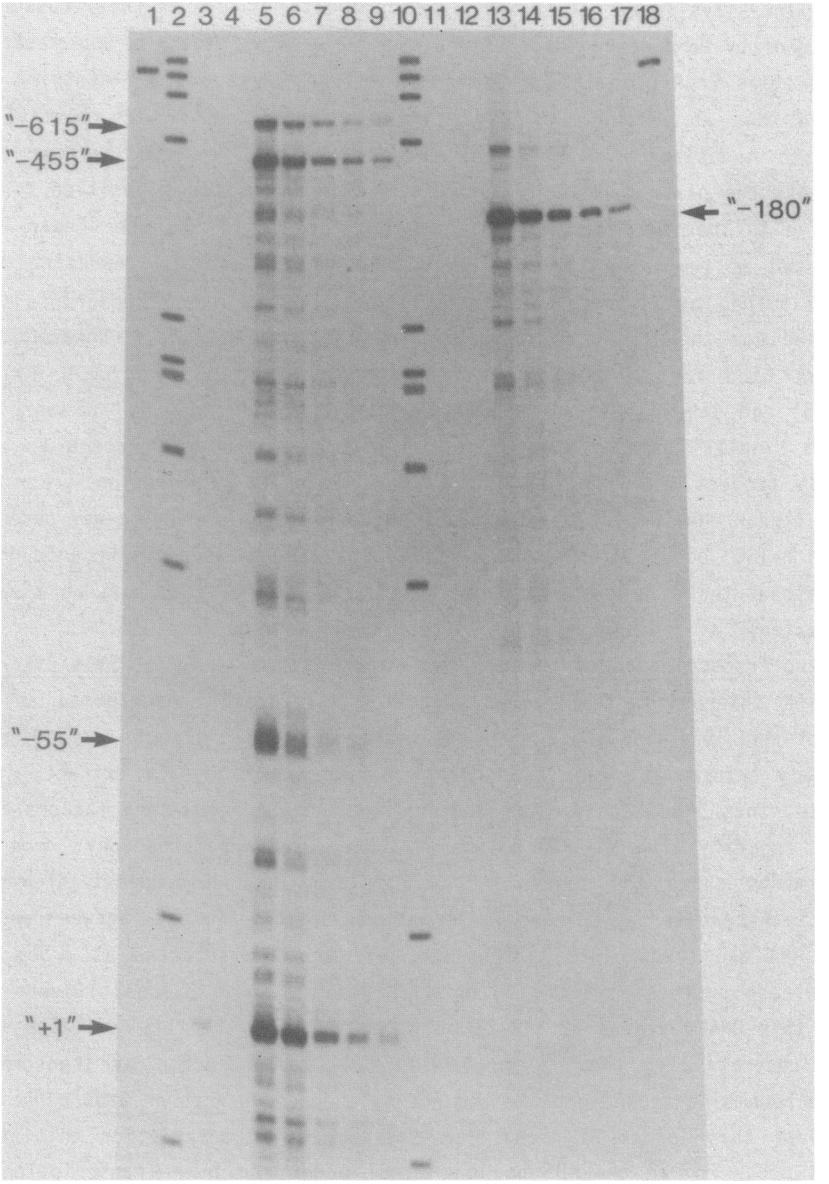


Figure 5. Efficiency of transcription in vitro of the rbcL and atpB genes. Five transcription reactions (40 µl) each contained 4 µl of spinach chloroplast high-salt extract and 6 µg/ml of pSoc802 DNA. The resulting in vitro synthesized RNAs were combined and serial dilutions of the RNAs were analyzed by S1 nuclease protection. The probe for S1 nuclease analysis was either the purified atpB-coding strand (lanes 3-9) or the purified rbcL-coding strand (lanes 11-17) of the 1134 bp MspI DNA from pSoc802 (15,000

cpm/assay). The amounts of in vitro RNAs analyzed were as follows: 40% (lane 5) 20% (lane 6), 10% (lane 7), 5% (lane 8), 2.5% (lane 9), 10% (lane 13), 5% (lane 14), 2.5% (lane 15), 1.25% (lane 16), 0.625% (lane 17). Three mock transcription reactions (40 µl) each contained 6 µg/ml of pSoc802 DNA and no high-salt extract. Of the combined reaction products two-thirds were analyzed for atpB transcripts (lane 3) and one-sixth were analyzed for rbcL transcripts (lane 11). Three other mock transcription reactions (40 µl) each contained 4 µl of high-salt extract and no exogenously added DNA. The reaction products were combined and two-thirds were analyzed for atpB transcripts (lane 4) and one-sixth were analyzed for rbcL transcripts (lane 12). Lane 1: 200 cpm of the purified atpB-coding DNA strand. Lanes 2 and 10: ØX174-HaeIII DNA fragments used as DNA size standards. Lane 18: 200 cpm of the purified rbcL-coding DNA strand. The positions of the 5' termini of in vitro synthesized atpB ("-615", "-455" and "+1") and rbcL ("-180") transcripts are indicated. Also shown is a presumed S1 artifact at position "-55" (see text).

script produced in vitro for each gene corresponds to the largest transcript seen in vivo, the "-180" and "-455" transcripts for the rbcL and atpB genes, respectively. In addition, there is an optimal template:extract ratio for in vitro transcription and high template concentration inhibits the production of stable RNA in vitro. Another interesting feature of the experiment shown in Figure 4 is that two minor atpB transcripts ("-615" and "-275") are also produced in vitro. The relative abundance of these two additional transcripts appears to vary according to the template:extract ratio. The 5' ends of these two transcripts are consistent with transcription initiation of the atpB gene from the rbcL start site ("-615") and from the 5' end region of the second major in vivo atpB mRNA ("-275"). However, these 5' end positions are at present tentative assignments based on the adjacent ØX174-HaeIII DNA size standards.

Efficiency of Transcription in Vitro

The S1 nuclease protection assay was used to estimate the efficiency of the in vitro transcription reaction (Figure 5). Two-fold serial dilutions of the in vitro RNAs were hybridized to each single-stranded DNA probe to ensure that a linear response was observed between increasing amounts of in vitro RNA and an increasing amount of S1 nuclease protection. As expected, the predominant in vitro rbcL transcript corresponds to the "-180" species. The transcription of the atpB gene is more complicated, however. In this experiment atpB "transcripts" are detected at positions "-615", "-455", "-55", and "+1". A "-615" atpB transcript is present in vivo as a very low abundance species (data not shown). The "-55" position on the non-coding strand of the atpB gene contains a run of twelve thymidines interrupted by a single adenosine (17). This "-55" signal is presumed to be an

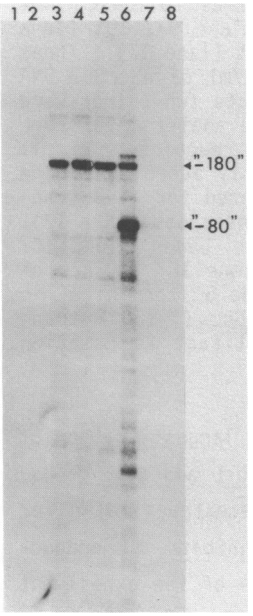


Figure 6. Transcription of the rbcL gene by chloroplast RNA polymerase is resistant to rifampicin. Transcription reactions (40 µl) contained either no high-salt extract (lane 2), 4 µl of the high-salt extract (lanes 1, 3-5) or 0.03 (BRL) units of E. coli RNA polymerase (lanes 6-8). Reactions contained either no exogenous DNA (lane 1) or pSoc802 DNA at 6µg/ml (lanes 2-8). All reactions were performed at 30°C. In this experiment 38% of the RNA from each reaction was analyzed by S1 nuclease protection assay, using a probe that is specific for in vitro synthesized RNA (see text). This probe was the 1134 nt (rbcL-coding) single-stranded MspI DNA of pSoc802 (15,000 cpm/assay). Transcription reaction conditions were as follows: Lane 1: spinach high-salt extract and no exogenous DNA. Lane 2: pSoc802 DNA and no enzyme. Lanes 3-5: pSoc802 DNA transcribed by the high-salt extract either in the absence (lane 3) or presence (20 µg/ml, lane 4; 120 µg/ml, lane 5) of rifampicin. Lanes 6-8: pSoc802 DNA transcribed by E. coli RNA polymerase either in the absence (lane 6) or presence (20 µg/ml, lane 7; 120 µg/ml, lane 8) of rifampicin. The positions of the 5' termini of the in vitro synthesized rbcL transcripts ("-180" and "-80") are indicated to the right of lane 8.

artifact due to S1 nuclease cleavage of an AT-rich sequence (22). A relatively large amount of in vitro transcription occurs at approximately the "+1" position (the beginning of the protein coding region of the atpB gene). This transcript is not found in vivo and is presumed to be an artifact of the in vitro transcription reaction. The intensity of the atpB background bands varies with different experiments (see Figure 4). The efficiency of atpB transcription may therefore be underestimated this experiment. The exposed regions were excised from the film and the silver grains were quantitated (15). Using the signal for the atpB "-455" transcript to normalize these data, the following relative intensities were obtained: atpB "-615", 0.4; atpB "-455", 1.0; atpB "+1", 1.2; and rbcL "-180", 6.4. The specific activity of the DNA probe was 3.08 x $10^6$ cpm/pmole (assuming 100% efficiency for the kinase reaction). The two DNA strands were examined on a strand separation gel to ensure that they had been labeled to an equal specific activity. The 200 cpm of probe shown in lanes 1 and 18 each represent 0.065 fmole of single-stranded DNA. Therefore, for each standard 40 µl transcription reaction the following amounts of transcripts were detected: atpB "-615", 0.07 fmole; atpB "-455", 0.16 fmoles; atpB "+1", 0.20 fmoles; rbcL "-180", 1.0 fmoles. These values correspond to one atpB "-455" transcript per 600 templates and one rbcL "-180" transcript per 100 templates.

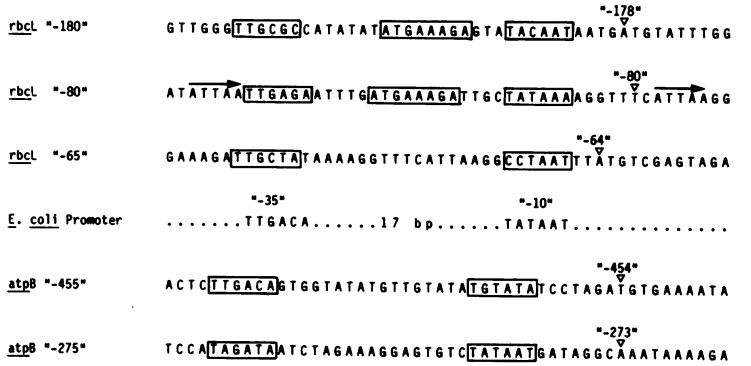


Figure 7. Comparison of DNA sequences that occur at the 5' termini of spinach rbcL and atpB transcripts. Two in vivo rbcL transcripts have been identified with 5' termini at positions "-178" (13, 16) and "-64" (13). In addition, E. coli RNA polymerase initiates transcription of this gene at approximately the "-80" position (see text). The DNA sequence surrounding each of these regions is shown above. Also shown is the E. coli consensus promoter sequence (5, 6) and the 5' terminal regions of the two largest in vivo atpB transcripts ("-454" and "-273"; 13). Chloroplast sequences homologous to the E. coli "-10" and "-35" promoter elements are enclosed in boxes. Also boxed is an octanucleotide present in both the "-180" and "-80" regions. Arrows indicate a 5 bp direct repeat (5'-ATTAA) that surrounds the "-80" promoter-like region.

## Transcription of the rbcL Gene is Resistant to Rifampicin

The DNA sequence homologies between the putative control regions of chloroplast genes with those of bacterial genes may reflect a similarity between chloroplast and bacterial RNA polymerases. One difference between the two enzymes, however, is their relative sensitivity to inhibition by rifampicin. The E. coli enzyme is inhibited by rifampicin while the chloroplast enzyme from higher plants is not (23, 24, 25). Early experiments (23, 24) assayed polymerase activity as total cpm incorporated into acid-insoluble material while a more recent experiment assayed the production of specific tRNAs in vitro in the presence and absence of rifampicin (25). As this report also discussed the possibility of multiple RNA polymerases in chloroplasts, we decided to examine if the potentially different enzyme responsible for mRNA synthesis was also insensitive to rifampicin. Transcription of the rbcL gene by the chloroplast high-salt extract was challenged with rifampicin (Figure 6). The products of in vitro transcription were essentially the same in either the absence (lane 3) or presence (20 µg/ml, lane 4; 120 µg/ml, lane 5) of this antibiotic. This result is consistent with earlier reports of rifampicin insensitivity. As a control in

this experiment, E. coli RNA polymerase was shown to be inhibited by both concentrations of rifampicin. However, it is interesting to note that in the absence of rifampicin (lane 6), the E. coli RNA polymerase initiates transcription of the rbcL gene at primarily two loci, the correct "-180" start site, and to an even greater extent at the rbcL "-80" position.

## DISCUSSION

In this paper we describe a system for in vitro transcription and analysis of chloroplast protein genes. This system uses either a crude high-salt extract or a DEAE-purified enzyme fraction. Both extracts will correctly initiate transcription of exogenously added chloroplast protein-coding genes. The high-salt extract contains endogenous nucleic acid and therefore the newly synthesized transcripts are assayed with a probe specific only for in vitro products (Figure 1). Most contaminating nucleic acids in the high-salt extract can be removed by simple DEAE column chromatography with a step salt gradient elution. However, a more complete separation of the endogenous nucleic acids from the chloroplast RNA polymerase is obtained by a linear salt gradient elution (J.E. Mullet, E.M. Orozco, Jr., and N.-H. Chua, unpublished observations).

The 5' termini of many chloroplast genes have been determined by primer extension and S1 nuclease protection analyses. However, only the rbcL gene has had its transcription initiation site determined unambiguously by in vitro "capping" analysis of the chloroplast mRNA (13). In spinach chloroplasts the "-180" position of the rbcL gene is the site of transcription initiation in vivo and the "-65" position is presumably a site of RNA cleavage. In the homologous in vitro system described here, transcription initiation of the rbcL gene occurs primarily at the correct "-180" in vivo start site and is independent of the vector DNA sequences. These results provide strong evidence that the same regulatory elements of the rbcL gene are recognized in vitro as in vivo. The atpB gene in spinach has four stable mRNAs with 5' termini at positions "-455", "-275", "-180" and "-100" (13). Although the primary in vivo transcript of the atpB gene has not been identified, in this report we have mapped the 5' terminus of the major in vitro atpB transcript to position "-455". In addition, a 135 bp fragment containing the corresponding position of the maize atpB gene has been cloned and is sufficient for transcription initiation in vitro using a maize chloroplast extract (L. Hanley-Bowdoin and N.-H. Chua, unpublished observations). These results indicate that this position is most likely the site of transcription

initiation of the atpB gene in vivo. However, it is interesting to note that under certain conditions (Figure 4 and unpublished observations) transcription initiation or RNA processing may occur in vitro at the atpB "-275" position. The DNA sequences preceding both the "-455" and the "-275" positions have some homology with the E. coli consensus promoter sequence (Figure 7).

Several other soluble chloroplast transcription systems have been described and the corresponding in vitro RNAs characterized. An active extract from maize has been used to transcribe the maize $tRNA^{His}$ and rbcL genes (26, 27). Transfer RNA genes are usually transcribed as precursor molecules (28), however, and as such the $tRNA^{His}$ transcript observed was likely a result of both transcription and processing in vitro. Jolly et al. (27) have observed in vitro synthesis of a maize rbcL transcript with its 5' terminus at the "-65" position, which is the 5' end of an in vivo maize rbcL transcript (29). However, the maize rbcL gene has recently been shown to also encode a larger "-300" transcript (13) and this "-300" region will support transcription initiation in vitro (30; L. Hanley-Bowdoin and N.-H. Chua, unpublished observations). Therefore, the original maize "-65" transcript produced in vitro may have been also the result of coupled transcription and processing. Gruissem et al. have developed a soluble system from spinach chloroplasts to examine the in vitro transcription and processing of various chloroplast tRNA genes (25). By use of Bal31 deleted templates, Gruissem et al. identified sequences between 56 and 85 bp upstream from the spinach $tRNA_m^{Met}$ gene that are required for optimal transcription of that gene (25). This deletion analysis appears to define the 5' end of the $tRNA_m^{Met}$ gene promoter, however, the precise in vivo and in vitro transcription initiation sites for the $tRNA_m^{Met}$ gene have not yet been determined.

A homologous transcription system from mustard chloroplasts has been reported recently (31). In that study transcription initiation in vitro was observed at the correct 5' end of the chloroplast psbA gene. (Only one chloroplast mRNA has been detected for the psbA gene in vivo.) In addition, a correctly-sized psbA "run-off" transcript was obtained from a linear (SmaI-digested) template. The mustard transcription system, however, is prepared quite differently from system described here. The mustard system may contain additional stromal factors not present in our spinach extracts. It will be interesting to determine if the mustard transcription system will correctly transcribe other linear DNA templates, and likewise if the spinach

transcription system described here will correctly transcribe a linear psbA gene. For the spinach rbcL gene, accurate transcription in our in vitro system requires a circular template. Linear templates invariably produce a large amount of nonspecific transcription initiations. Possibly the site at which the template has been linearized provides an "entry site" for the chloroplast RNA polymerase as has been reported to occur for RNA polymerase II in vitro (32). The polymerase could then initiate transcription at this site as well as other downstream sequences which fortuitously resemble the chloroplast promoter sequence. We are currently investigating whether superhelicity of the circular DNA template has any effect on rbcL gene transcription in vitro.

We have made a rough estimation of the efficiency of the in vitro transcription system. For the experiment described in Figure 5, one atpB "-455" transcript is produced per 600 templates per hour and one rbcL "-180" transcript is produced per 100 templates per hour. However, this is most likely a conservative estimate because of the following assumptions: (1) 100% efficiency of the kinase reaction, which determines the specific activity of the DNA probe; (2) No loss of RNA in the manipulations (e.g. phenol:$CHCl_3$:isoamyl alcohol extractions) following the in vitro transcription reactions; and (3) 100% efficiency of the S1 nuclease protection assay. For the atpB transcript in particular, a high AT base content may contribute to a decrease in the "-455" signal due to artifactual S1 cleavage.

We have used the in vitro system to demonstrate that spinach chloroplast RNA polymerase is resistant to high concentrations of rifampicin (Figure 6), as has been reported previously by others. In addition, in this experiment E. coli RNA polymerase recognized at least two regions in the spinach rbcL gene as promoters, producing a minor transcript at "-180" and a major transcript at "-80". These two transcripts are also detected in vivo in JM103 cells (L. Hanley-Bowdoin, E.M. Orozco, and N.-H. Chua, manuscript submitted). Figure 7 compares the chloroplast DNA sequence preceding the "-180" and "-80" positions to the E. coli consensus promoter sequence. Also included is the DNA sequence preceding the "-65" position, which is the 5' end of a major transcript in vivo but not in vitro. Both the "-180" and "-80" regions contain prokaryotic-like "-10" and "-35" sequences as well as a conserved octanucleotide (5'-ATGAAAGA) between these two elements. These two regions may have originally arisen by a sequence duplication event. Although the E. coli RNA polymerase will initiate transcription at both of

these regions, the chloroplast RNA polymerase does not recognize the "-80" region as a transcription start site. These results emphasize that the consensus promoter sequence recognized by E. coli RNA polymerase is not sufficient to constitute a chloroplast promoter region in our in vitro transcription system. However, the possibility exists that both the "-180" and "-80" regions may function as chloroplast promoters in vivo under some environmental or cell growth conditions. It will be interesting to determine if the rbcL "-80" region will promote transcription in vitro on templates in which the "-180" sequence has been removed.

We are currently using this in vitro transcription system to define promoter and regulatory regions of various chloroplast protein genes. These transcriptionally active extracts should also be useful for the isolation of factors that modulate chloroplast gene activities.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM30726 and a grant from the Monsanto Company. J.E.M. was supported by NIH postdoctoral fellowship GM08114. We are grateful to L. Hanley-Bowdoin and C.A. Edwards for helpful discussions. 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.

*Present address: Agricultural Research Service, U.S. Department of Agriculture and Departments of Agronomy and Plant Biology, University of Illinois, Urbana, IL 61801, USA
+Present address: Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843, USA

REFERENCES

1. Bohnert, H.J., Crouse, E.J. and Schmitt, J.M. (1982) in Encyclopedia of Plant Physiology, Pathier, B. and Boulter, D. Eds., Vol. 14B, pp. 475-530, Springer Verlag, New York.
2. Whitfeld, P.R. and Bottomley, W. (1983) Ann. Rev. Plant Physiol. 34, 279-310.
3. Scott, N.S. and Possingham, J.V. (1980) J. Exp. Bot. 31, 1081-1092.
4. Berry-Lowe, S.L., McKnight, T.D., Shah, D.M. and Meagher, R.B. (1982) J. Mol. Appl. Genet. 1, 483-498.
5. Rosenberg, M. and Court, D., (1979) Ann. Rev. Genet. 13, 319-353.
6. Hawley, D.K. and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237-2255.
7. Bartlett, S.G., Grossman, A.R. and Chua, N.-H. (1982) in Methods in Chloroplast Molecular Biology, Edelman, M., Hallick, R.B. and Chua, N.-H. Eds., pp. 1081-1091, Elsevier Biomedical Press, Amsterdam, The Netherlands.
8. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1522.
9. Cohen, S.N. and Miller, C.A. (1970) J. Mol. Biol. 50, 671-687. 10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor New York.
11. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
12. Tullis, R.H. and Rubin, H. (1980) Anal. Biochem. 107, 260-264.

13. Mullet, J.E., Orozco, E.M., Jr. and Chua, N.-H. (1984) Plant Mol. Biol. (in press).
14. Weaver, R.F. and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
15. Suissa, M. (1983) Anal. Biochem. 133, 511-514.
16. Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R. (1981) Nucleic Acids Res. 9, 3251-3270.
17. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) Proc. Natl. Acad. Sci. USA 79, 6260-6264.
18. Erion, J.L., Tarnowski, J., Peacock, S., Caldwell, P., Redfield, B., Brot, N. and Weissbach, H. (1983) Plant Mol. Biol. 2, 279-290.
19. Viera, J. and Messing, J. (1982) Gene 19, 259-268.
20. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Cell 18, 469-484.
21. Manley, J.L., Fire, A., Cano, A., Sharp, P. and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3855-3859.
22. Hansen, U., Tenen, D.G., Livington, D.M. and Sharp, P.A. (1981) Cell 27, 603-612.
23. Polya, G.M. and Jagendorf, A.T. (1971) Arch. Biochem. Biophys. 146, 635-648.
24. Bottomley, W., Smith, H.J. and Bogorad. L. (1971) Proc. Natl. Acad. Sci. USA. 68, 2412-2416.
25. Gruissem, W., Greenberg, B.M., Zurawski, G., Prescott, D.M. and Hallick, R.B. (1983) Cell 35, 815-828.
26. Schwartz, Z., Jolly, S.O., Steinmetz, A.A. and Bogorad, L. (1981) Proc. Natl. Acad. Sci. USA 78, 3423-3427.
27. Jolly, S.O., McIntosh, L., Link, G. and Bogorad, L. (1981) Proc. Natl. Acad. Sci. USA 78, 6821-6825.
28. Soll, D., Abelson, J.N. and Schimmel, P.R. Eds., (1980) Transfer RNA: Biological Aspects. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
29. McIntosh, L., Poulsen, C. and Bogorad, L. (1980) Nature 288, 556-560.
30. Crossland, L.D., Rodermel, S.R. and Bogorad, L. (1984) Proc. Natl. Acad. Sci. USA 81, 4060-4064.
31. Link, G. (1984) EMBO J. 3, 1697-1704.
32. Grosschedl, R. and Birnstiel, M.L. (1982) Proc. Natl. Acad. Sci. USA 79, 297-301.