*Plant Molecular Biology* **21**: 1069–1076, 1993.



# Co-transcription pattern of an introgressed operon in the maize chloroplast genome comprising four ATP synthase subunit genes and the ribosomal *rps2*

D.J. Stahl[1,3], S.R. Rodermel[2,4], L. Bogorad[2] and A.R. Subramanian[1,*]

[1] *Max Planck Institute for Molecular Genetics, Ihnestrasse 73, W-1000 Berlin 33, FRG (*author for correspondence);* [2] *Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA, USA; Present adresses:* [3] *Planta, Postfach 146, W-3352 Einbeck, FRG;* [4] *Department of Botany, Iowa State University, Ames, IA 50011, USA*

Received 1 July 1992; accepted in revised form 10 January 1993

*Key words:* maize (*Zea mays*), chloroplast operon, chloroplast transcription, ribosomal protein genes, thylakoid ATP synthase

## Abstract

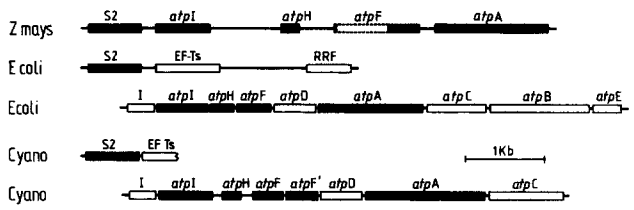
Several examples of the introduction of a gene from one gene complex into another (introgression) are found when chloroplast RP gene clusters are compared to those in *Escherichia coli* or cyanobacteria. Here we describe the transcript pattern of one such cluster from maize (*Zea mays*) that includes the genes for 4 subunits of the thylakoid ATP synthase (*atpI, H, F, A*) and the *rps2* gene. Twelve transcript species covering the size range from 7000 to 800 nt were identified in RNA isolated from dark-grown and greening maize seedlings, and several of them were characterized by reverse transcription analysis. A major species of 6200 nt, with its 5′ end at 181 nt upstream of the initiating ATG of *rps2*, contained the transcripts of all the 5 genes. Two further sets of transcripts having their 5′ ends ca. 120 and 50 nt upstream of the initiation codons of the *atpI* and *atpH* genes were also identified. Thus, this plastid gene cluster in maize is functionally organized as an operon with additional regulatory features to allow for increased accumulation of mRNAs for the thylakoid components.

## Introduction

The gene for chloroplast ribosomal protein (RP) S2 in the maize chloroplast genome occurs in a cluster that includes four genes for the subunits of the ATP synthase complex of the photosynthetic apparatus [1]. Since translational and photosynthetic systems [2] differ so considerably in the quantitative requirements of their protein components, complex regulatory features would be necessary if RP genes are to be integrated with photosynthetic genes into common transcription units.

In *Escherichia coli*, RP-S2 is encoded in a gene cluster with the genes for the elongation factor EF-Ts and the ribosome release factor RRF [3, 4], whereas all the subunits of the ATP synthase are encoded in a single *atp* (old name *unc*) operon [5]. These two loci are at 4 min and 84 min on the *E. coli* chromosomal map [6], i.e. at a distance of about $1 \times 10^6$ bp from each other, and they are divergently transcribed. In two cyano-

bacterial species so far examined (*Spirulina platensis* and *Synechococcus* sp. [7, 8], the *rps2* and *atp* genes are localized in unlinked clusters. These situations are as depicted below (the genes under discussion shown filled):




If chloroplasts evolved from free-living, photosynthetic bacteria, then these genes for two functionally distinct multiprotein complexes, viz. the ribosome and the ATP synthase, have been brought together during the course of chloroplast evolution. If they also form a functional transcription unit, then this process has involved not only gene transposition but also promoter remodelling. In this paper we show that the introgressed genes of the *rps2-atpIHFA* cluster in the maize chloroplast genome are co-transcribed and therefore they form a functional operon. The transcript pattern of the operon and the 5′- end analyses of the transcripts (both in dark and in light) are described.

## Materials and methods

### *Preparation of maize RNA*

(*Zea mays* seeds (FR9cmsTR37; Illinois Foundation Seeds Corp, Champaign, IL) were germinated and grown in the dark on moist vermiculite at 25 °C for 7 days in an environmental chamber and were either directly processed under green safelight (Schott, Mainz [etiolated]), or exposed to light (5000 lux) for 16 h before processing (greening). Leaves were excised and quick-frozen in liquid $N_2$ and RNA was extracted by the guanidinium thiocyanate/CsCl procedure [9] with some modifications. In brief, 4 g lots of frozen leaves were ground in liquid $N_2$ and vortexed in 20 ml of 4 M guanidinium thiocyanate, 0.1 M 2-mercaptoethanol, 0.025 M sodium citrate pH 7.0, 0.5% sodium lauroylsarcosine. After filtration (Miracloth) and centrifugation (3000 × *g*, 15 min), 0.5 g/ml of CsCl was added to the supernatant and the RNA was sedimented through a cushion of 5.7 M CsCl and 0.1 M $Na_2EDTA$ at 130000 × *g* for 18 h. The RNA was dissolved in 10 mM Tris-HCl, pH 7.6, 5 mM $Na_2EDTA$ and 0.1% SDS, extracted with chloroform/butanol (4:1), precipitated with ethanol, and dissolved in sterile $H_2O$; the concentration was determined from $UV_{260nm}$ absorption [10].

### *Northern blot analysis*

RNA samples (15 μg/lane) were denatured in 10 mM sodium phosphate (pH 6.5) and 50% DMSO with glyoxal (1 M, 50 °C, 1 h), electrophoresed in a 1.1% agarose gel in 10 mM sodium phosphate, pH 6.5, and blotted onto a nylon membrane in the presence of 20 × SSC [10]. The filter was prehybridized overnight at 50 °C in 5 × SSC, 50 mM sodium phosphate pH 6.5, 8 × Denhardt's solution and 0.25 mg/ml denatured calf thymus DNA and then hybridized overnight with [$^{32}$P]-labelled (random-primer procedure, Amersham) DNA probes. The blots were washed 2 × 20 min at 60 °C in 0.1 × SSC, 0.1% SDS and autoradiographed. For quantitation, the autoradiograms (at several exposure times) were scanned on a Quickscan Densitometer (Desaga) and the areas under peaks were determined by integration.

### *Primer extension analysis*

Each radioactively-labelled primer was mixed with maize RNA (40–100 pg primer to 20 μg RNA in 12 μl of 0.4 M NaCl, 45 mM PIPES, pH 6.0); and the sample was denatured for 1 min at 90 °C and then annealed at 60 °C for 6 h. Primer extension was carried out with 10 units of AMV reverse transcriptase (Promega), in the

presence of 0.5 M Tris-HCl pH 8.3, 6 mM $MgCl_2$, 40 mM KCl, 1 mM DTT, 0.1 mg/ml bovine serum albumin and 0.5 mM each of dATP, dGTP, dCTP and dTTP (final volume 100 µl), at 42 °C for 1 h. The reaction products were precipitated with ethanol and separated either on a 5% polyacrylamide gel with labelled molecular size markers, or on an 8% polyacrylamide gel with [$^{32}$P]dATP-labelled products of a dideoxy sequencing reaction [11] using the same primer, and autoradiographed.

The primers used in these experiments were restriction fragments of the *rps2-atpIHFA* gene cluster [1, 12] isolated from agarose gels. They include: 1) 135 nt long *Eco* RI-*Hinf* I fragment (Pr-2, position 881–1015 [1]); 2) 45 nt long *Eco* RI-*Ssp* I (Pr-3, position 1855–1899 [1]); 3) 110 nt long *Ssp* I-*Sca* I (position 1900–2009 [1], data not shown); 4) 64 nt long *Sca* I-*Hinf* I (Pr-4, position 2010–2073 [1]); 5) 68 nt long *Bsp* HI-*Apa* I (Pr-5, position 4–71 [12]). In addition, a 35-mer deoxyoligonucleotide (Pr-1, position 851–885 [1]) was synthesized (Applied Biosystems DNA synthesizer Model 380A) and purified by reversed-phase HPLC and used. The restriction fragment primers were labelled with [$^{32}$P] by the fill-in reaction and isolated by electrophoresis through denaturing polyacrylamide gels [10]; the synthetic oligomer was labelled with [$\gamma$-$^{32}$P]ATP in the presence of T4 polynucleotide kinase [10].

## Results and discussion

### *Transcript analysis*

The *rps2-atpIHFA* gene cluster is located in the single-copy region of maize chloroplast DNA at map positions 72–78.5 kb [1, 13]. It is embedded in the 28 kb inversion in cereal plastid genomes [14] and lies immediately downstream of the *rpoC2* gene and upstream of the *rps14-trnfM-trnR* cluster (Fig. 1). We hybridized DNA probes (see Materials and methods for details) specific to the *rpoC2*, *rps2*, *atpA*, *atpF*, *atpH* and *atpI* genes and the intron in *atpF* (illustrated in Fig. 2B) to northern blots of RNA isolated from greening and etiolated plants. The results for all 8 probes are summarized in Fig. 2A (probes B and B′ gave identical results and therefore are shown as B/B′) and are diagrammed schematically in Fig. 2B.

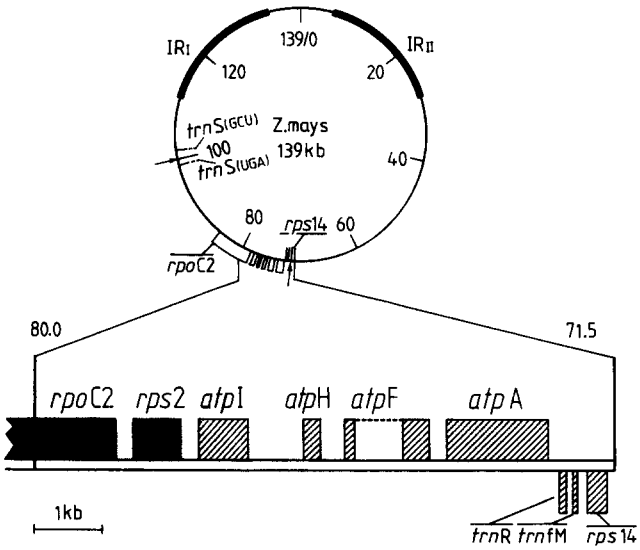



*Fig. 1.* Location and orientation of the *rps2-atpIHFA* gene cluster in the maize chloroplast genome [1, 13]. The termini of the 28 kb long inversion are indicated by the two arrows in the large single-copy region.

A total of 12 RNA bands ranging in size from 800 nt to 7000 nt were detected by the 8 probes (marked by arrowheads in Fig. 2A). With the exception of the two smallest (800 nt and 1000 nt), all the others were polycistronic transcripts. There was one major transcript (6200 nt) which hybridized to all 8 probes, i.e. it comprised transcripts of *rps2* and the 4 ATP synthase genes. This result showed that *rps2* is cotranscribed with the ATP synthase genes in maize chloroplast DNA.

Five transcript species (4900 nt, 3100 nt, 2700 nt, 2500 nt and a weak signal of 1800 nt; Fig. 2) hybridized to probes of *atpI* (B/B′) but did not hybridize to the *rps2* probe. Primer extension analyses (see below) indicated that these transcripts have a common 5′ end located in the spacer between *rps2* and *atpI*. Another group of 4 transcripts (3200, 1500, 1000 and 800 nt) hybridized to the *atpH* probe (Probe C) but did not hybridize to probes upstream of *atpI* (i.e. Probes A, B/B′). Primer extension experiments suggested

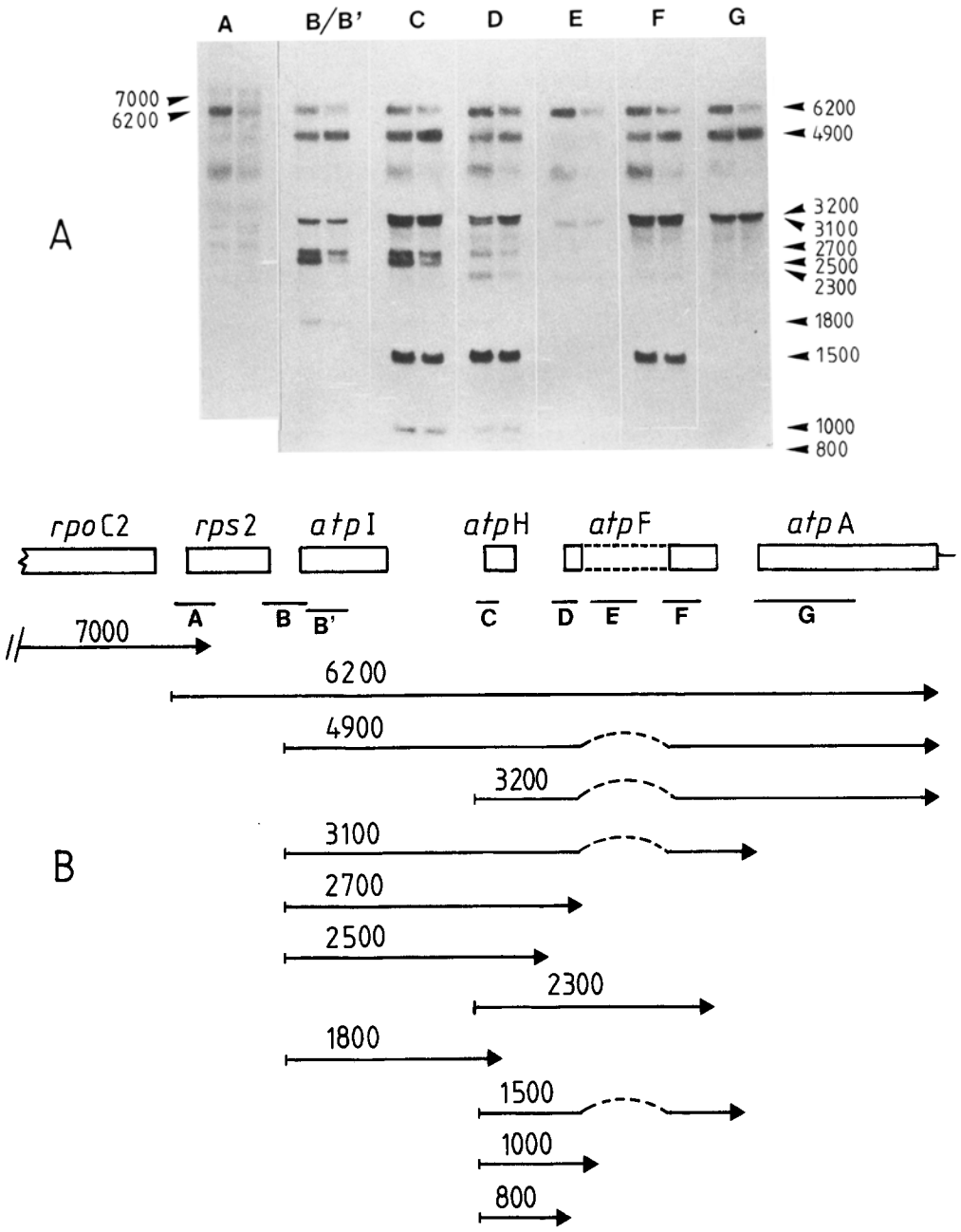



*Fig. 2.* Northern blot analysis of the transcripts from *rps2-atpIHFA* gene cluster. The positions of the 8 probes (A to G) in the various genes are shown in the middle. The hybridization pattern (with two RNA preparations from greening [left] and etiolated [right] plants) is shown in panel A (a diffuse signal at ca. 4200 nt seen with several of the probes probably could arise from nonspecific binding to the rRNAs banding in this region; it is therefore not considered). Numbers designate the nucleotide lengths of the indicated RNA bands. The detected transcripts are shown schematically in panel B. Dots represent the spliced out intron in *atpF*.

that these transcripts have a common 5′ end located upstream of the *atpH* gene.

One of the probes (E in Fig. 2) is derived from the intron in *atpF*, the only intron-containing gene of this cluster. This probe detected the 6200 nt transcript with an intensity similar to that given by the other probes. Thus the large, 6200 nt long transcript species would be the primary transcript of the entire *rps2-atpIHFA* operon.

In these northern hybridizations, each probe was hybridized to two RNA preparations designated 'greening' and 'etiolated' (left and right lanes in Fig. 2A). The results show an influence of light on the amounts of the 6200 nt, 2700 nt and 2500 nt transcripts but little effect on the levels of the other transcripts. These results are consistent with those of Rodermel and Bogorad [15] showing that the multiple transcripts from this region of the maize chloroplast genome increase differentially in amount during greening.

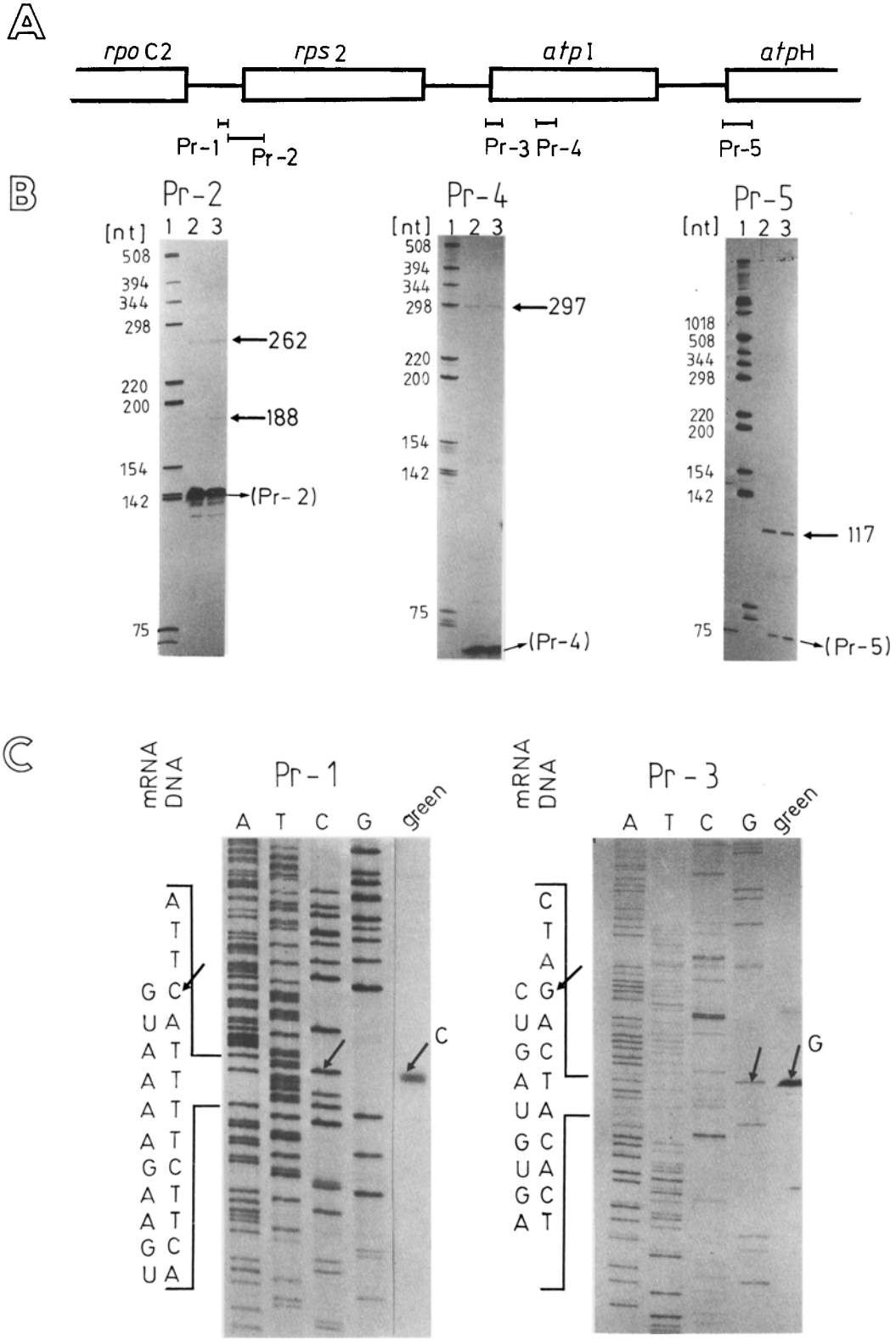
*The 5′ ends of transcripts*

Primer extension of transcripts with reverse transcriptase was performed using 6 single-stranded primers from the upstream regions of *rps2*, *atpI*, and *atpH* genes (see Materials and methods). The results are summarized in Fig. 3. The upstream *rps2* primer (Pr-2; 135 nt, *Eco* RI-*Hinf* I) revealed two potential start sites (bands marked 262 and 188 in Fig. 3B). Since the 3′ end of this 135 nt long primer is exactly 52 bp upstream of the initiating ATG of *rps2* [1], these two putative transcription start signals would correspond, respectively, to initiations at 181 and 107 bp upstream of the ATG.

In the published nucleotide sequence of the *rps2-atpIHFA* gene cluster [1] a chloroplast promoter motif, 'TTGTAC – ' 17 bp ' – TTTAAT', occurs 184–212 bp upstream of the *rps2* initiator ATG (and 80–109 bp downstream of the *rpoC2* terminator TAA). A transcript from this putative promoter (see Fig. 4) would correspond to the stronger signal (262 nt), in Fig. 3B. This inference is supported by the high-resolution primer extension experiment using the synthetic 35-mer deoxyoligonucleotide (Pr-1). From the sequencing gel of this experiment the starting base of the transcript was determined to be G, corresponding to the transcribed DNA strand sequence CATTTTC... (Fig. 3C), 3 nucleotides downstream from the −10 element. The presence of a second promoter (corresponding to the 188 nt signal) could not be confirmed either by examining the nucleotide sequence (Fig. 4) for a promoter motif or from additional experiments (data not shown). The shorter signal therefore probably represents a premature reverse transcriptase termination.

Primer extension analysis of the intergenic region between *rps2* and *atpI* was performed with three primers (45 nt *Eco* RI-*Ssp* I, Pr-3, 110 nt *Ssp* I-*Sca* I (data not shown); and 64 nt *Sca* I-*Hinf* I, Pr-4): all gave similar results. The experiment with the 64 nt primer (Pr-4) is shown in Fig. 3B. The 297 nt signal corresponds to a 5′ end about 120 bp upstream of the initiating ATG of *atpI*. A chloroplast promoter motif, 'TTGATA – ' 20 bp ' – TATATT', occurs 126–157 bp upstream of the ATG of *atpI* in the published sequence [1]. A high resolution primer extension experiment using Pr-3 (Fig. 3C) showed that this transcript species may start with the base C, corresponding to the transcribed DNA strand sequence GACTAC..., 6 nucleotides downstream from the putative −10 element. This 5′ end would be common to the five transcripts 4900, 3100, 2700, 2500 and 1800 (Fig. 2B). Whether it really represents the start point of a common promoter or only a specific mRNA processing site remains to be determined.

Primer extension analysis for the transcripts specific to *atpH* and its downstream genes *atpF* and *atpA* was performed with a 68 nt (*Bsp* HI-*Apa* I, Pr-5) fragment. One strong signal, 117 nt in length, was detected (Fig. 3b). It corresponds to a 5′ end nucleotide 50 bp upstream of the initiating ATG of *atpH*. There is a chloroplast promoter-like sequence, 'TAGAGC – ' 17 bp ' – TAATTT', 55–83 bp upstream of this initiating codon of *atpH* [1]. The mapped 5′ end, GTTGA... (data not shown), is located 5 bp downstream of the −10 element and will be com-


A
rpoC2
rps2
atpI
atpH
Pr-1
Pr-2
Pr-3
Pr-4
Pr-5
B
Pr-2
[nt] 1 2 3
508
394
344
298
220
200
154
142
75
262
188
(Pr-2)
Pr-4
[nt] 1 2 3
508
394
344
298
220
200
154
142
75
297
(Pr-4)
Pr-5
[nt] 1 2 3
1018
508
344
298
220
200
154
142
75
117
(Pr-5)
C
mRNA
DNA
Pr-1
A T C G green
A T T G C U A T A T A T A G C A T A T G C U A
C
mRNA
DNA
Pr-3
A T C G green
C T A C G U A G C A T U A G C U G C A T
G

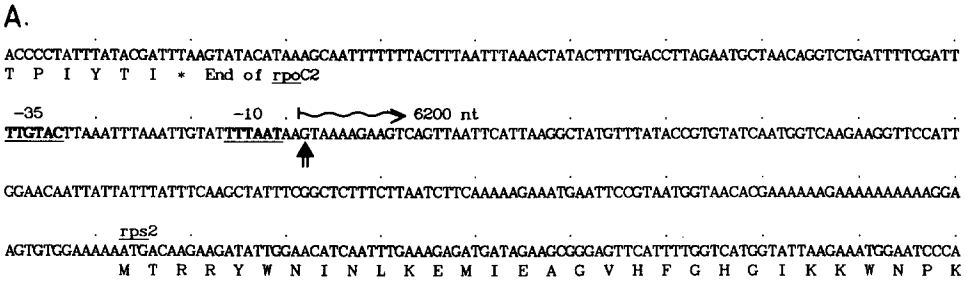



*Fig. 4.* The promoter motif upstream of the *rps2* gene [1]. The nucleotide (G) complementary to the identified 5′ end of the 6400 nt RNA (Fig. 2) is shown by an arrow.

mon to the five transcripts, 3200, 2300, 1500, 1000 and 800 nt.

Further experiments, for example 5′ capping of mRNA, are necessary to establish the actual functionality of these three promoter motifs. Interestingly, such motifs are not present in the previously reported chloroplast *rps2-atpIHFA* sequences from pea and spinach [16].

As with the northern experiments, the primer-extension analyses were also performed with RNA preparations from both etiolated and greening maize seedlings. The two RNA preparations gave essentially the same result and thus there was no indication for any additional light-activated promoters or processing pathways.

### *Quantitation of transcript amounts*

The relative amounts of the transcripts for the 5 genes of the *rps2-atpIHFA* operon were determined by quantitative densitometry (see Materials and methods) taking the 6200 nt long transcript, which includes all the 5 genes as the standard. The level of *rps2* transcripts was considerably lower than that of the ATP synthase genes, the ratio *rps2*:*atpI*:*atpH*:*atpF*:*atpA* being 1.0:7.5:16.5:6:4.5. The abundance levels of the ATP synthase mRNAs to some degree reflected the stoichiometry of the corresponding polypeptides in the ATP synthase [2], e.g. the most abundant polypeptide in the $F_1 \cdot F_0$ complex (10–12 copies) is subunit c, the translation product of the most abundant *atpH* mRNA.

Considered together, the data in this paper are consistent with the hypothesis that accumulation of the individual proteins encoded by the *rps2-atpIHFA* gene cluster is primarily a function of transcription initiation (for a recent report on the role of transcription rate in barley chloroplasts, see [17]) but further experiments are needed to confirm it and to rule out the possibility that protein accumulation is also a function of differential mRNA processing, stability and other post-transcriptional effects (such as translational and post-translational controls). These data complement Haley and Bogorad's [18] observation that two other complex transcription units in the maize chloroplast genome contain internal promoters that are regulated differentially by light, and that protein accumulation is largely a function of the molar levels of the various mRNAs complementary to given genes of these complexes.

Our results are similar in several respects to the transcript analyses carried out earlier on the *rps2-atpIHFA* gene cluster in pea and spinach [16]. This cluster in these two dicot plants generates multiple transcripts that fall into several families with 5′ ends upstream of the *rps2*, *atpI*, and *atpH*

*Fig. 3.* Determination of the 5′ ends of the RNA species by primer extension with reverse transcriptase. A. Positions and sizes of the used primers (Pr-1 to Pr-5). B. The lengths of three primers after chain extension. C. Identification of the 5′-end nucleotides of *rps2* and *atpI* gene transcripts by high-resolution analysis after sequencing reaction. The results with 'greening RNA' are shown; 'etiolated RNA' gave the same results.

genes. In contrast to maize however, the largest transcripts in these systems did not include the complete *rps2* coding region. This paper thus presents the first evidence for cotranscription of the entire *rps2-atpIHFA* cluster as a single, detectable polycistronic mRNA. The data suggest that the regulation of the *rps2-atpIHFA* operon in the monocots and dicots shows similarities as well as differences indicating that the process of promoter remodelling following the formation of a fused gene cluster (another example: *Z. mays* L23-II operon with truncated S10, *spc* and alpha, see [19, 20]) is an ongoing process in the evolution of higher-plant chloroplast genomes.

## References


1. Stahl D, Rodermel SR, Subramanian AR, Bogorad L: Nucleotide sequence of a 3.46 kb region of maize chloroplast DNA containing the gene cluster *rpoC2-rpS2-atpI-atpH*. Nucl Acids Res 18: 3073–3074 (1990).
2. Bogorad L, Vasil IK (eds): The Photosynthetic Apparatus: Molecular Biology and Operation. Academic Press, New York (1991).
3. An G, Bendiak DS, Mamelak LA, Friesen JD: Organization and nucleotide sequence of a new ribosomal operon in *Escherichia coli* containing the genes for ribosomal protein S2 and elongation factor Ts. Nucl Acids Res 9: 4163–4172 (1981).
4. Ichikawa S, Kaji A: Molecular cloning and expression of ribosome release factor. J Biol Chem 264: 20054–20059 (1989).
5. Walker JE, Gray NJ, Saraste M, Eberle AN: DNA sequence around the *Escherichia coli unc* operon. Biochem J 224: 799–815 (1984).
6. Bachmann BJ: Linkage map of *Escherichia coli* K-12, edition 8. Microbiol Rev 54: 130–197 (1990).
7. Sanangelantoni AM, Calogero RC, Buttarelli FR, Gualerzi CO, Tiboni O: Organization and nucleotide sequence of the genes for ribosomal protein S2 and elongation factor Ts in *Spirulina platensis*. FEMS Microbiol Lett 66: 141–146 (1990).
8. Cozens AL, Walker JE: The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. J Mol Biol 194: 359–383 (1987).
9. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294–5299 (1979).
10. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1982).
11. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
12. Rodermel SR, Bogorad L: Molecular evolutions and nucleotide sequences of the maize plastid genes for the α subunit of $CF_1$ (*atpA*) and the proteolipid subunit of $CF_0$ (*atpH*). Genetics 116: 127–139 (1987).
13. Larrinua IM, Muskavitch KMT, Gubbins EJ, Bogorad L: A detailed restriction map of the *Zea mays* plastid genome. Plant Mol Biol 2: 129–140 (1983).
14. Sugiura M: The chloroplast chromosomes in land plants. Annu Rev Cell Biol 16: 293–301 (1989).
15. Rodermel S, Bogorad L: Maize plastid photogenes: mapping and photoregulation of transcript levels during light-induced development. J Cell Biol 100: 463–476 (1985).
16. Hudson GS, Mason JG, Holton TA, Koller B, Cox GB, Whitfeld PR, Bottomley W: A gene cluster in the spinach and pea chloroplast genome encoding one $CF_1$ and three $CF_0$ subunits of the $H^+$-ATP synthase complex and the ribosomal protein S2. J Mol Biol 196: 283–298 (1987).
17. Rapp JC, Baumgartner BJ, Mullet J: Quantitative analysis of transcription and RNA levels of 15 barley chloroplast genes. J Biol Chem 267: 21404–21411 (1992).
18. Haley J, Bogorad L: Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. Plant Cell 2: 232–333 (1990).
19. Ruf M, Kössel H: Structure and expression of the gene coding for the α subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. Nucl Acids Res 16: 5741–5754 (1988).
20. Weglöhner W, Subramanian AR: Nucleotide sequence of maize chloroplast *rpl*32: completing the apparent set of plastid ribosomal protein genes and their tentative operon organization. Plant Mol Biol, in press (1993).