Mol Gen Genet (1987) 209:427–431

MGG

Rapid splicing and stepwise processing of a transcript from the *psbB* operon in tobacco chloroplasts: Determination of the intron sites in *petB* and *petD*

Minoru Tanaka, Junichi Obokata*, Julapark Chunwongse, Kazuo Shinozaki, and Masahiro Sugiura**
Center for Gene Research, Nagoya University, Chikusa, Nagoya 464, Japan

Summary. Expression of the *psbB* gene cluster in tobacco chloroplasts has been studied. This cluster contains the genes for the 51 kDa chlorophyll *a* apoprotein (*psbB*) and the 10 kDa phosphoprotein (*psbH*) of photosystem II, and cytochrome *b*6 (*petB*) and subunit IV (*petD*) of the cytochrome *b*/*f* complex in this order. Northern blot hybridization and reverse transcription analyses have revealed that *petB* and *petD* contain single introns and the *psbB* gene cluster is transcribed as a single polycistronic unit. The primary transcript seems to be spliced very rapidly and then processed into several small RNA species. The exact splice sites have been located by cDNA sequencing. The transcriptional initiation site of the *psbB* operon has been determined by S1 mapping with *in vitro* capped chloroplast RNA. The stepwise processing of chloroplast RNA precursors is discussed.

Key words: Tobacco chloroplast – *psbB* operon – Intron – Splicing – Polycistronic transcription

Introduction

Chloroplasts contain unique transcriptional and translational systems different from the nuclear-cytoplasmic systems. The chloroplast systems have been shown to be similar to prokaryotic systems rather than eukaryotic systems (Dyer 1984). Chloroplast genomes contain many gene clusters which have been shown to be cotranscribed; *rrn* operons, *atp* clusters, *trn* clusters, *psb* clusters and ribosomal protein gene clusters (e.g. Strittmatter and Kössel 1984; Shinozaki et al. 1983; Deno et al. 1984; Westhoff et al. 1985; Cozens et al. 1986; Westhoff 1985; Ohme et al. 1985; Tanaka et al. 1986). Multiple RNA species have often been observed for one given gene in these gene clusters. The complex patterns of the RNA species have been thought to be the results of multiple RNA processing of the primary transcripts.

Barkan et al. (1986) have observed striking alterations in the abundance and size of RNA species in *psbB* gene cluster when chloroplast RNA extracted from a photosynthetic mutant of maize (*hcf-38*) was compared with that from the wild type. This finding suggests that RNA processing in chloroplasts may be involved in a post-transcriptional regulation of chloroplast gene expression.

The *psbB* gene cluster is known to contain at least four genes, the genes for the 51 kDa chlorophyll *a* apoprotein (*psbB*), the 10 kDa phosphoprotein (*psbH*, formerly designated *psbF*, Shinozaki et al. 1986b), cytochrome *b*6 (*petB*) and subunit IV (*petD*) in this order, and has been suggested to be polycistronically transcribed (Morris and Herrmann 1984; Heinemeyer et al. 1984; Phillips and Gray 1984). Here we show that the tobacco *petB* and *petD* genes contain single introns and that the transcript from the *psbB* gene cluster is rapidly spliced in tobacco chloroplasts.

Materials and methods

Total tobacco chloroplast RNA was prepared from young leaves grown in a growth chamber at 25° C for 12 h illumination as described (Shinozaki and Sugiura 1982). To remove DNA contamination, 1 mg of the RNA preparation was treated at 37° C for 1 h with 10 units of bovine pancreatic DNaseI in the presence of 140 units of human placenta ribonuclease inhibitor (HPRI) (Takara Shuzou Co., Ltd.). After phenol treatment, RNA was precipitated with ethanol.

Reverse transcription analysis. A primer was hybridized to 100 μg of total tobacco chloroplast RNA in 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 6 mM $MgCl_2$ and 140 units of HPRI at a temperature gradually decreasing from 65° C to 42 °C for 2 h. cDNA was synthesized at 42° C for 30 min in a reaction mixture containing 25 mM Tris-HCl (pH 8.3), 25 mM KCl, 3 mM $MgCl_2$, 0.1 mM each of 4 dNTPs, 0.2 mM of either ddNTP, 5 units of AMV reverse transcriptase (Boehringer Mannheim), 10 mM DTT and 50 μg/μl actinomycin D (Zaita et al. 1987). After incubation this reaction mixture was treated with 1 μl 1 N NaOH at 100° C for 3 min and neutralized with 1 μl of 1 N HCl.

S1 mapping was performed as described (Shinozaki and Sugiura 1982). To label 5'terminal triphosphates of chloroplast RNA, 100 μg of tobacco chloroplast RNA was incubated with [α-^{32}P]GTP in the presence of 1.8 units of guanydyltransferase (Bethesda Research Laboratories) in 300 μl of reaction buffer containing 50 mM Tris-HCl (pH 8.0), 1.25 mM $MgCl_2$, 6 mM KCl and 2.5 mM DTT (Crossland et al. 1984). Labelled (capped) RNA was precipitated with

* *Present address:* Research Center for Molecular Genetics, Hokkaido University, Sapporo 060, Japan
** *Present address:* Faculty of Agriculture, Kasetsart University, Phaholyothin Rd, Bangkok, 10900, Thailand
Offprint requests to: M. Sugiura

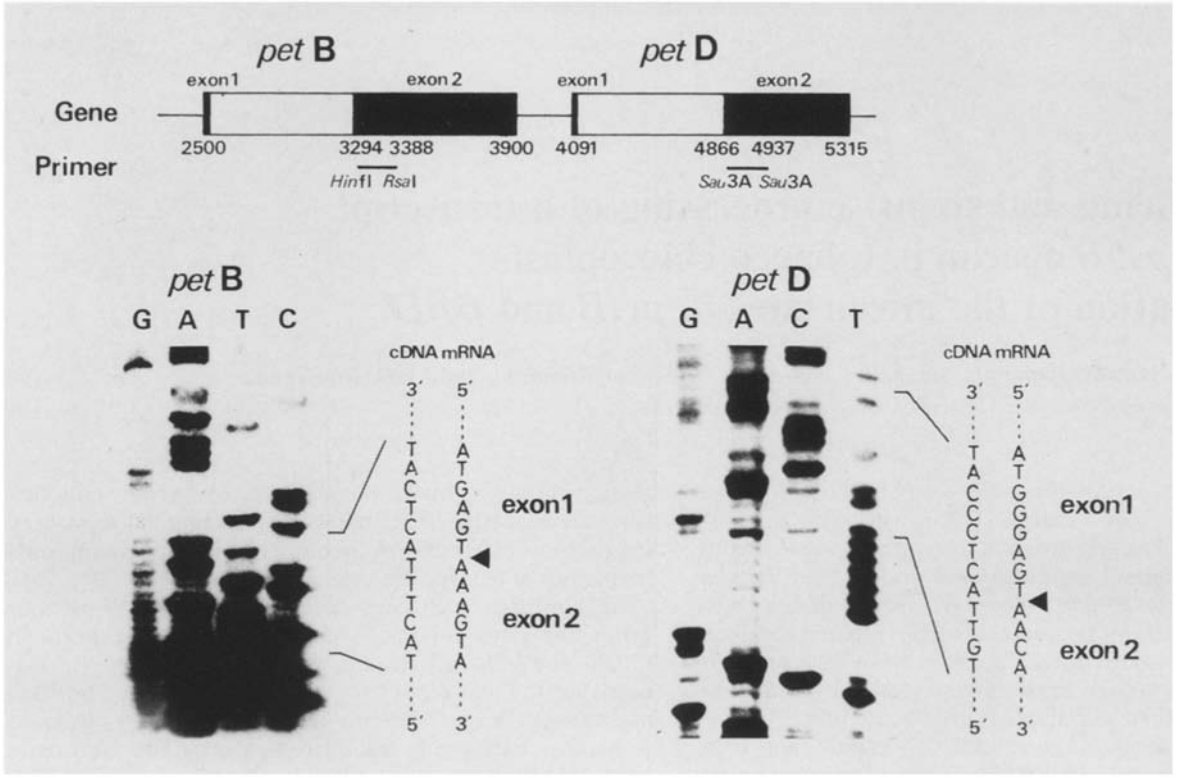

Fig. 1. Determination of the splice sites of a precursor RNA from *petB* and *petD*. The *upper* part represents *petB* and *petD* structures. Exons are shown as *filled boxes* and introns as *open boxes*. *Bars* indicate primers used for reverse transcription. The numbering starts from the first nucleotide of *psbB*. The 1095 bp *Xho*I/*Pst*I fragment containing the *petB* exon 2 and the 72 bp *Sau*3A fragment containing the *petD* exon 2 were cloned into M13mp18 or M13mp19 vectors. Coding strands were synthesized with [α-^{32}P]dCTP by Klenow fragment of DNA polymerase I and cut with *Hin*fI/*Rsa*I and *Sau*3A as primers for *petB* and *petD*, respectively. The *lower* part shows autoradiograms of cDNA sequencing gels. *Triangles* indicate the splice sites

ethanol after addition of 2 μg of 572 bp *Nsi*I DNA fragment (see Fig. 4a) and subjected to S1 nuclease digestion. DNA segments protected by the capped chloroplast RNA were electrophoresed in a 4% polyacrylamide gel containing 8 M urea.

Results

Introns of petB *and* petD

The nucleotide sequence of the tobacco *psbB* gene cluster (*psbB–psbH–petB–petD*) has been reported (Shinozaki et al. 1986b; see Fig. 3). Computer analysis of the nucleotide sequences of *petB* and *petD* suggested the presence of long introns in the *petB* and *petD* genes. To prove the existence of the introns, reverse transcription analysis was performed to detect the corresponding spliced RNA species. Primers were prepared from the coding strands of the putative exon 2 in *petB* and *petD* (see Fig. 1) and hybridized to total tobacco chloroplast RNA. cDNAs were synthesized from the primers using AMV reverse transcriptase in the presence of dideoxyribonucleoside triphosphates.

RNA species in which the expected exon 1 sequences were joined with the expected exon 2 sequences from *petB* and *petD* were detected as shown in the sequencing ladders (Fig. 1). These ladders indicate the exact splice sites of the precursor RNA. We therefore concluded that there were introns of 753 bp in *petB* and 742 bp in *petD*. No significant open reading frames were found in either strand of these introns. The intron boundary sequences are (5')GTGTGAC—GCCTATCTCAAT(3') in *petB* and (5')GTGTGAC—ACCTATCCCAAT(3') in *petD* (Fig. 2), which match the conserved boundary sequences of chloroplast group III introns (Shinozaki et al. 1986a). The sequence ladders in Fig. 1 show little divergence after the splice sites, indicating that the primary transcript is scarce.

Only two amino acids residues can be deduced from exon 1 in *petB* and a sequence similar to the prokaryotic ribosome binding site, GGAG, was found 2 bp upstream of this exon (Fig. 2). In the case of *petD*, there are two possible initiation codons. We tentatively assigned the proximal ATG as the initiation codon because the distal GTG codon is not conserved in the *petD* sequence of spinach (Heinemeyer et al. 1984). Hence the *petD* exon 1 is only 8 bp (Fig. 2): a similar short exon has also been found in the gene for ribosomal protein CL16 in chloroplast DNA (Tanaka et al. 1986; Posno et al. 1986).

Expression of the psbB *operon*

Northern blot hybridization was performed to analyse how this *psbB* gene cluster is expressed in tobacco chloroplasts. As shown in Fig. 3, all probes for the *psbB, psbH, petB* and *petD* coding regions hybridized to a common RNA band of 3.9 kb while the 3.9 kb RNA band could not be detected by probes derived from introns (1.5 kb in total) of *petB* and *petD*. Based on the length of *psbB* gene cluster (5.4 kb) together with the reverse transcription analysis de-

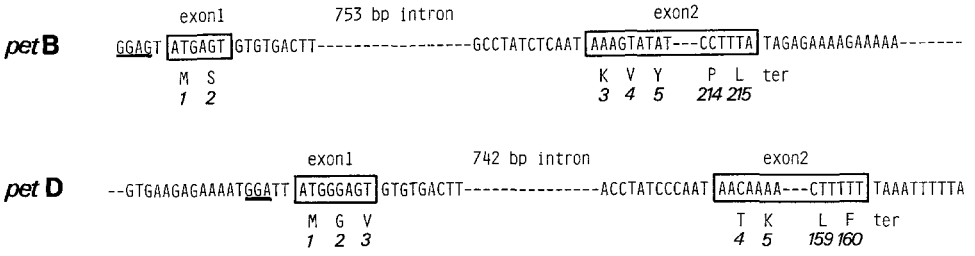

Fig. 2. Portions of the nucleotide sequences of *petB* and *petD*. Exons are *boxed* and the predicted amino acid sequences are shown *below* the DNA sequences with residue numbers. Putative ribosomal binding sites are *underlined*

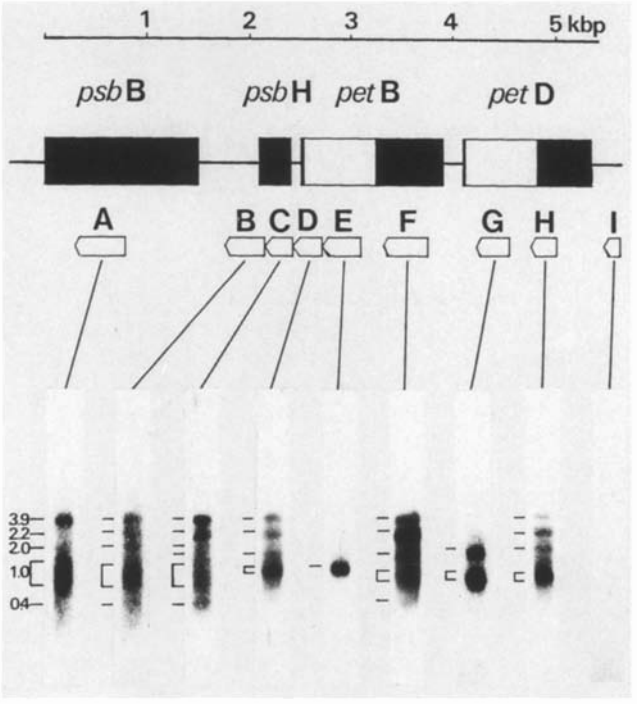

Fig. 3. Detection of transcripts from the *psbB* operon. The *upper* part shows the organization of the *psbB* operon. Exons are shown as *filled boxes* and introns as *open boxes*. The numbering starts from the first nucleotide of *psbB*. *psbB*, positions 1–1527 (1527 bp); *psbH*, positions 2149–2370 (222 bp); *petB*, positions 2500–3900 (1401 bp, a 753 bp intron, positions 2506–3258); *petD*, positions 4091–5315 (1225 bp, a 742 bp intron, positions 4099–4840). DNA fragments indicated by *open arrows* (A–I) were cloned into M13mp18 or M13mp19 vectors and labelled with [α-^{32}P]dCTP by Klenow fragment as probes. A, 509 bp *Eco*RI/*Alu*I; B, 383 bp *Xmn*I/*Eco*RI; C, 200 bp *Eco*RI/*Sac*I; D, 270 bp *Sac*I/*Eco*RI; E, 236 bp *Eco*RI/*Eco*RI; F, 585 bp *Xho*I/*Sna*BI; G, 449 bp *Eco*RI/*Eco*RI; H, 233 bp *Eco*RI/*Ssp*I; I, 155 bp *Eco*RI/*Ssp*I. The *lower* part shows the autoradiograms of RNA blot hybridization. Total tobacco chloroplast RNA was sizefractionated on 1.0% agarose gels, transferred to nylon membranes and hybridized with the above probes. Numbers in the leftmost column are kb

scribed above, the 3.9 kb RNA species is most likely to be a spliced RNA molecule encompassing the entire *psbB* gene cluster with the *psbB* gene cluster being transcribed as a unit (operon). The intron sequences derived from *petB* and *petD* seems to be spliced out very rapidly from the primary transcript, or removed while RNA is being transcribed, in tobacco chloroplasts under our growth conditions. Since several discrete RNA bands of less than 3.9 kb were detected by probes for the *psbB*, *psbH*, *petB* and *petD* coding regions, the spliced 3.9 kb RNA species is probably subjected further to stepwise processing, the rate of which seems slower than that of splicing.

The RNA bands (>1 kb) detected by probes for the *petB* and *petD* introns are apparently longer than the *petB* and *petD* introns, 753 bp and 742 bp, respectively. This suggests that the intron sequences excised do not remain as linear forms but become unusual forms such as lariats which have a retarded mobility. We detected at least two RNA bands with the *petD* intron probe while only one band with the *petB* intron probe. This difference remains to be explained.

Transcriptional initiation and termination of the psbB *operon*

S1 mapping was performed to determine the site of transcriptional initiation. Total tobacco chloroplast RNA was hybridized to a 441 bp *Ssp*I/*Nsi*I fragment labelled with [^{32}P] at its 5' end (Fig. 4a) and then digested with S1 nuclease. Two protected bands of 248–250 bp and of 124–127 bp were detected (Fig. 4b). These two bands correspond to positions 175–177 bp and 51–54 bp, respectively, upstream of the *psbB* coding region in the DNA sequence (Fig. 4c). Similarly two sites have been reported in the spinach *psbB* gene cluster (Westhoff 1985).

To examine whether these two sites are the initiation points, we performed an additional S1 mapping using the capped chloroplast RNA. Total tobacco chloroplast RNA was labelled at the 5'terminal triphosphates with guanydyltransferase in the presence of [α-^{32}P]GTP and hybridized to the 572 bp *Nsi*I DNA fragment (see Fig. 4a). The capped RNA/DNA hybrid was digested with S1 nuclease and subjected to size fractionation in parallel with a size marker and the capped RNA (no S1 nuclease treatment). We obtained only one capped RNA band of 248–252 bp and concluded that the distal site, 175–177 bp upstream of *psbB*, is the initiation site; the proximal site, 51–54 bp upstream of *psbB*, turned out to be a processing site (Fig. 4c). There are sequences similar to the Pribnow box and −35 region in front of the initiation site. This region is very conserved between tobacco and spinach *psbB* operons (Fig. 4c). No RNA bands were detected for the region downstream of *petD* by Northern blot analysis when a 155 bp *Eco*RI/*Ssp*I DNA fragment (67 bp downstream of *petD*) was used as probe (Fig. 3, lane 1). The termination site is probably located within the 67 bp flanking region at the 3' end of *petD*.

Discussion

The 3.9 kb RNA species is the longest detectable RNA transcribed from the *psbB* operon of tobacco grown under our growth conditions. This 3.9 kb RNA species contains the entire coding sequences of the operon and lacks the intron

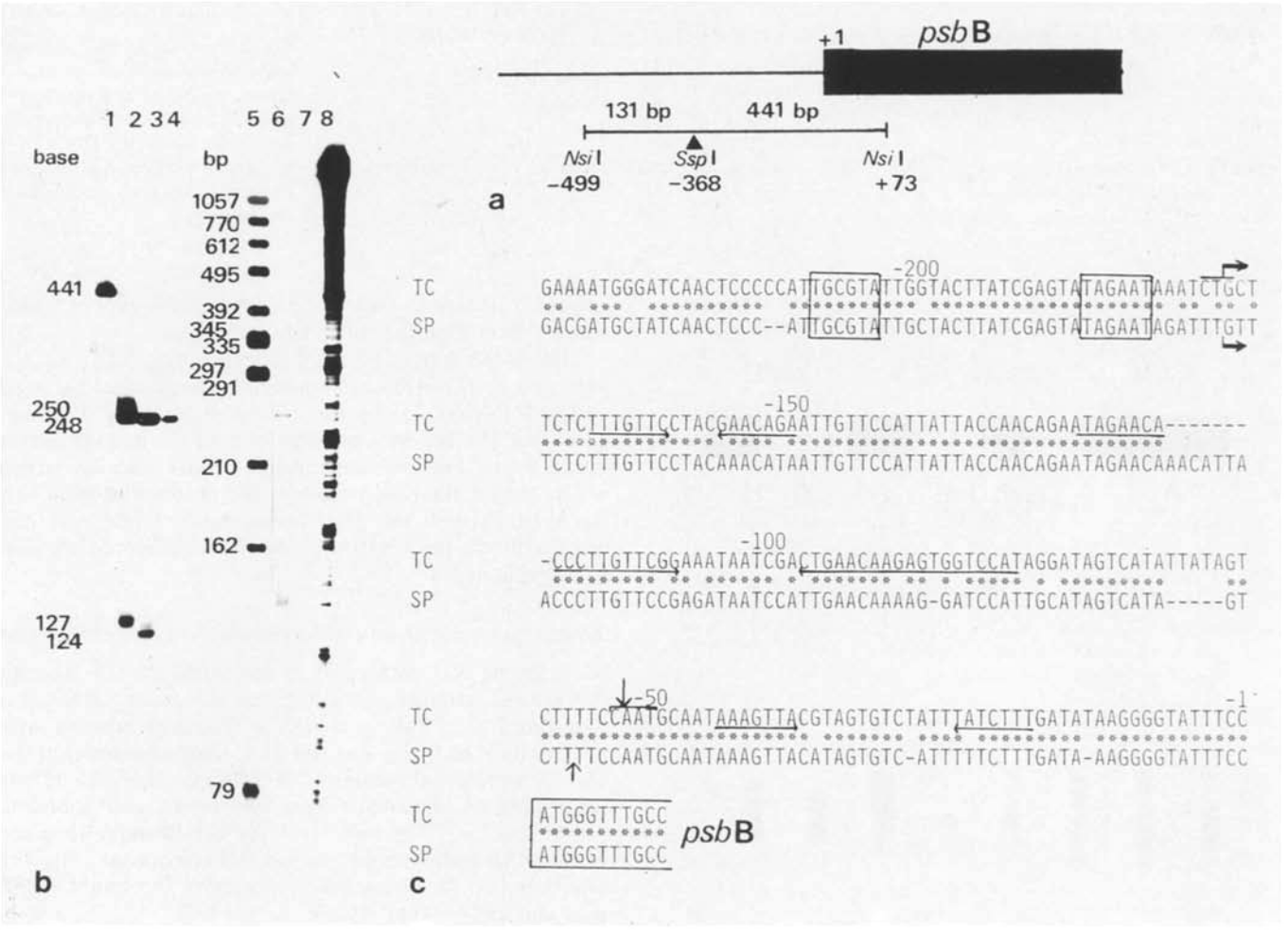

Fig. 4a–c. Determination of the transcriptional initiation site of the *psbB* operon by S1 mapping analyses. **a** Location of DNA fragments used for S1 mapping analysis. The numbering starts from the first nucleotide of *psbB*. **b** Autoradiograms of S1 nuclease protected fragments. The 572 bp *Nsi*I DNA fragment was end-labelled with [γ-^{32}P]ATP by polynucleotide kinase, cut with *Ssp*I and size fractionated in a 5% polyacrylamide gel, and the 441 bp *Ssp*I/*Nsi*I fragment was eluted from the gel. Total tobacco chloroplast RNA (17 µg) was hybridized with the heat-denatured [^{32}P]fragment. Lane 1, 441 bp *Ssp*I/*Nsi*I DNA fragment; Lanes 2, 3 and 4, S1 nuclease-protected DNA fragments (5 units, 10 units and 20 units, respectively). Total chloroplast RNA (100 µg) was labelled (capped) with [α-^{32}P]GTP by guanydyltransferase and hybridized with the 572 bp heat-denatured *Nsi*I DNA fragment. Lane 5, end-labelled *Hinc*II-digest of ϕX174 DNA as size markers; Lanes 6 and 7, S1 nuclease protected capped RNA (20 units and 50 units, respectively); Lane 8, the capped total chloroplast RNA (no treatment). **c** Comparison of promoter regions of the *psbB* operon between tobacco (TC) and spinach (SP). Identical nucleotides are shown by *asterisks*. Sequences similar to the −10 region and −35 region are *boxed*. Inverted repeats are *underlined*. Transcriptional starting sites are shown by *horizontal arrows* and the processing site by *vertical arrows*

sequences of *petB* and *petD*. The reverse transcription analysis showed that spliced RNA species lacking the intron sequences is extremely dominant over the primary and unspliced RNA species; this allowed us to determine the splice sites of the precursor RNA.

Initiation of transcription of the *psbB* operon was determined to be 175–177 bp upstream of the *psbB* coding region by S1 mapping using the capped chloroplast RNA. No other transcriptional initiation sites were detected between *psbB* and *psbH* (data not shown), indicating that the 3.9 kb RNA species is not a mixture of RNAs that start from the regions upstream of *psbB* and between *psbB* and *psbH* (the length from *psbH* to *petD* is about 3.9 kb). No transcripts were detected from a region 67 bp or further downstream of the 3'end of *petD*, indicating that the transcription ends within the 67 bp 3' flanking region of *petD*. We therefore conclude that the tobacco *psbB, psbH, petB,* and *petD* genes are transcribed polycistronically as a single unit. This is compatible with that reported in the spinach *psbB* gene cluster (Westhoff 1985). As we could not detect a potential primary transcript, which should be at least 5.4 kb long, covering this operon, the intron sequences in the pre-mRNA molecule seem to be rapidly removed or splicing initiates during transcription. Rapid splicing has also been observed in *trans*-splicing of transcripts from the divided gene for ribosomal protein CS12 (5'-*rps*12 and 3'-*rps*12, Zaita et al. 1987) and in *cis*-splicing of transcripts from *ndhA* and *ndhB* (Matsubayashi et al. 1987) in tobacco chloroplasts.

No transcripts starting in the 621 bp region between *psbB* and *psbH* were detected, suggesting that the smaller RNA species detected by Northern blot analysis are generated from the 3.9 kb RNA species. One of the processing sites was located by S1 mapping 51–54 bp upstream from

the *psbB* coding region and resulting in removal of the 5' leader sequence of about 125 bases from the primary transcript. The complex pattern of RNA species raises the possibility that post-transcriptional regulation may operate in the expression of *psbB* operon by the cleavage of RNA species. One of the possibilities is that the 3.9 kb RNA species is processed stepwise in a fixed order into functional monocistronic mRNAs. An alternative possibility is that the spliced 3.9 kb RNA species is a tetracistronic mRNA for the four proteins and the smaller RNA species are degradation products.

Acknowledgements. This work was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture and a grant from the Toray Science Foundation.

References

Barkan A, Miles D, Taylor WC (1986) Chloroplast gene expression in nuclear, photosynthetic mutants of maize. EMBO J 5:1421–1427

Cozens AL, Walker JE, Phillips AL, Huttly AK, Gray JC (1986) A sixth subunit of ATP synthase, an F_O component, is encoded in the pea chloroplast genome. EMBO J 5:217–222

Crossland LD, Rodermel SR, Bogorad L (1984) Single gene for the large subunit of ribulose-1,5,-bisphosphate carboxylase in maize yields two differentially regulated mRNAs. Proc Natl Acad Sci USA 81:4060–4064

Deno H, Shinozaki K, Sugiura M (1984) Structure and transcription pattern of a tobacco chloroplast gene coding for subunit III of proton-translocating ATPase. Gene 32:195–201

Dyer TA (1984) The chloroplast genome: Its nature and role in development. In: Baker NR, Barber J (eds) Chloroplast biogenesis. Elsevier, Amsterdam, pp 23–69

Heinemeyer W, Alt J, Herrmann RG (1984) Nucleotide sequence of the clustered genes for apocytochrome *b*6 and subunit 4 of the cytochrome *b/f* complex in the spinach plastid chromosome. Curr Genet 8:543–549

Matsubayashi T, Wakasugi T, Shinozaki K, Yamaguchi-Shinozaki K, Zaita N, Hidaka T, Meng BY, Ohto C, Tanaka M, Kato A, Maruyama T, Sugiura M (1987) Mol Gen Genet (in press)

Morris J, Herrmann RG (1984) Nucleotide sequence of the gene for the P_{680} chlorophyll *a* apoprotein of the photosystem II reaction center from spinach. Nucleic Acids Res 12:2837–2850

Ohme M, Kamogashira T, Shinozaki K, Sugiura M (1985) Structure and cotranscription of tobacco chloroplast genes for $tRNA^{Glu}_{UUC}$, $tRNA^{Tyr}_{GUA}$ and $tRNA^{Asp}_{GUC}$. Nucleic Acids Res 13:1045–1056

Phillips AL, Gray JC (1984) Location and nucleotide sequence on the gene for the 15.2 kDa polypeptide of the cytochrome *b-f* complex from pea chloroplasts. Mol Gen Genet 194:477–484

Posno M, Vliet A, Groot GSP (1986) The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to *E. coli* ribosomal protein L16 is split by a large intron near its 5'end: structure and expression. Nucleic Acids Res 14:3181–3195

Shinozaki K, Sugiura M (1982) The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Gene 20:91–102

Shinozaki K, Deno H, Kato A, Sugiura M (1983) Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase. Gene 24:147–155

Shinozaki K, Deno H, Sugita M, Kuramitsu S, Sugiura M (1986a) Intron in the gene for the ribosomal protein S16 of tobacco chloroplast and its conserved boundary sequences. Mol Gen Genet 202:1–5

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Y-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986b) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Strittmatter G, Kössel H (1984) Cotranscription and processing of 23 S, 4.5 S and 5 S rRNA in chloroplasts from *Zea mays*. Nucleic Acids Res 12:7633–7647

Tanaka M, Wakasugi T, Sugita M, Shinozaki K, Sugiura M (1986) Genes for the eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): Similarity to the S10 and *spc* operons of *Escherichia coli*. Proc Natl Acad Sci USA 83:6030–6034

Westhoff P (1985) Transcription of the gene encoding the 51 kd chlorophyll *a*-apoprotein of the photosystem II reaction centre from spinach. Mol Gen Genet 201:115–123

Westhoff P, Alt J, Nelson N, Herrmann RG (1985) Genes and transcripts for the ATP synthase CF_O subunits I and II from spinach thylakoid membranes. Mol Gen Genet 199:290–299

Zaita N, Torazawa K, Shinozaki K, Sugiura M (1987) *Trans* splicing in vivo: joining of transcripts from the "divided" gene for ribosomal protein S12 in the chloroplasts of tobacco. FEBS Lett 210:153–156

Communicated by R.B. Goldberg

Received April 30, 1987 / May 5, 1987