Volume 17 Number 23 1989 Nucleic Acids Research

An additional promoter within the protein-coding region of the *psb*D-*psb*C gene cluster in tobacco chloroplast DNA

Wen Biao Yao, Bing Yuan Meng, Minoru Tanaka and Masahiro Sugiura

Center for Gene Research, Nagoya University, Nagoya 464-01, Japan

Received September 20, 1989; Revised and Accepted November 10, 1989 EMBL accession no. Z00044

ABSTRACT

Transcription of the *psb*D-*psb*C gene cluster in tobacco chloroplasts has been studied. This cluster contains in linear sequence the overlapping genes encoding the D2 and 43 kDa proteins of Photosystem II (*psb*D and *psb*C, respectively), and ORF62. Eight major transcripts ranging from 1.5 to 4.4 kb were detected by northern blot analysis. S1 mapping experiments revealed that these multiple transcripts comprise five distinct 5'ends whose precise positions were further determined by primer extension analysis. Two of the five 5'ends were determined to be the transcriptional initiation sites by *in vitro* capping assays: the main site is located 905 bp upstream from the ATG codon of *psb*D and the additional site is 194 bp upstream from the first ATG codon of *psb*C. The latter site and the preceding prokaryotic promoter motif are within the protein-coding region of *psb*D. The 3'ends of transcripts were determined by S1 mapping.

INTRODUCTION

Photosystem II (PSII) is one of the four supermolecular complexes of chloroplast thylakoid membranes which are involved in the light reaction of photosynthesis (1, 2). The PSII complex functions as a water : plastoquinone oxidoreductase and contains approximately 20 different polypeptides (2), of which at least 11 components are encoded on the chloroplast genome (3). D2 and 43 kDa proteins are encoded on the chloroplast genome (*psb*D and *psb*C, respectively). The *psb*D and *psb*C genes overlap by 50 bp and are translated in two different reading frames in all higher plants so far examined (e.g. 4, 5, 6).

The transcription of *psb*D and *psb*C were reported for several plant species. At least six major RNA species with sizes ranging from 1.6 to 4.6 kb were found in spinach (4) and the two overlapping genes were suggested to be cotranscribed (4, 5). In dark-grown barley at least six transcripts ranging in size from 1.7 to 5.7 kb were detected, the four larger transcripts hybridized to *psb*D and *psb*C while the two smaller transcripts hybridized only to *psb*C (7). Illumination of dark-grown barley caused decrease in the six transcripts found in dark-grown barley and the accumulation of two different transcripts of 3.2 and 4.0 kb which hybridized to both *psb*D and *psb*C (8). Four transcripts of 2.6 to 5.0 kb and several smaller transcripts were detected in pea and of them a 1.5 kb transcript strongly hybridized to *psb*C (9). However there have been no reports on precise transcriptional initiation sites of the two overlapping genes and putative promoter sequences have simply been noted upstream of *psb*D (4, 5, 6).

To obtain more information on the transcriptional organization of chloroplast overlapping genes, we have analyzed transcripts from the tobacco *psb*D-*psb*C cluster and found two transcriptional initiation sites for the gene cluster; one is located upstream of *psb*D and

the other is within the coding region of *psb*D. To our knowledge, this is the first report to demonstrate a transcriptional initiation site within a protein-coding region in the chloroplast genome.

MATERIALS AND METHODS

DNA fragments

Recombinant plasmid pTB30 containing the 11.1 kbp *Bam*HI partial fragment (Ba19-9a-9b) has been described (10). Appropriate short DNA pieces derived from the plasmid DNA were either subcloned into pUC118 vector or extracted from gels after electrophoresis. Single-strand DNAs (coding-strands) were obtained from subclones in pUC118 with helper phage M13K07. Oligodeoxyribonucleotides were prepared by a DNA synthesizer (Applied Biosystems 380A) and labeled at their 5'ends with ^{32}P.

Total chloroplast RNA

Tobacco (*Nicotiana tabacum* var. Bright Yellow 4) was grown in a greenhouse for 2−3 months and total chloroplast RNA was prepared as previously described (11).

Northern blot hybridization

Total tobacco chloroplast RNA was separated in 1.1% agarose/6% formaldehyde gels and blotted onto a nylon membrane sheet and subsequently hybridized with [5'^{32}P] oligonucleotide probes according to the BioRad's instruction manual except that our prehybridization solution contained 1% SDS instead of 7%.

S1 mapping

S1 mapping was performed as described (12). 3'end labeling was carried out using a DNA 3'end labeling kit (Boehringer Mannheim) and [α^{32} P]ddATP.

Primer extension.

This was carried out using AMV reverse transcriptase as described (13).

In vitro capping

The 5'terminal triphosphates of tobacco chloroplast RNA were labeled with [α^{32} P]GTP and guanylyltransferase (BRL) as previously described (11).

RESULTS

Detection of psbD-psbC transcripts

To detect transcripts from the *psb*D-*psb*C cluster, northern blot analysis was performed using [5'^{32}P] oligonucleotide probes (probes A−K). As shown in Fig. 1, probes B−J hybridized to two common RNA bands of 4.4 and 3.6 kb in length. The 4.4 kb RNA was not detected with probes A (*trn*T) and K (*trn*G), indicating that the 4.4 kb transcript starts after *trn*T and ends before *trn*G. Probes B−J, which span 4.1 kbp, hybridized also to a 3.6 kb RNA band, suggesting that there are two RNA species which have the same size of 3.6 kb but differ in their ends. We assigned tentatively the 5'end of one 3.6 kb RNA right after *trn*T and the 3'end of another 3.6 kb RNA between ORF62 and *trn*G as shown in Fig. 1 (designated as 3.6a and 3.6b, respectively). Probes D−J, spanning 3.2 kbp, detected an additional common RNA band of 2.6 kb. This also suggests that the 2.6 kb RNA band contains two transcripts which differ in their ends (designated as 2.6a and 2.6b). A 2.4 kb RNA band was detected with probes G−J but not with probe F, suggesting that the 2.4 kb RNA has its 5'end different from the 2.6 kb RNA (2.6b). Finally, 1.7 kb and 1.5 kb RNA bands were found with probes F−H and probes G−H, respectively, and their tentative positions are shown in Fig. 1. Several other faint RNA bands were also found between 1.5 kb and 0.3 kb RNA bands, and these may be processed

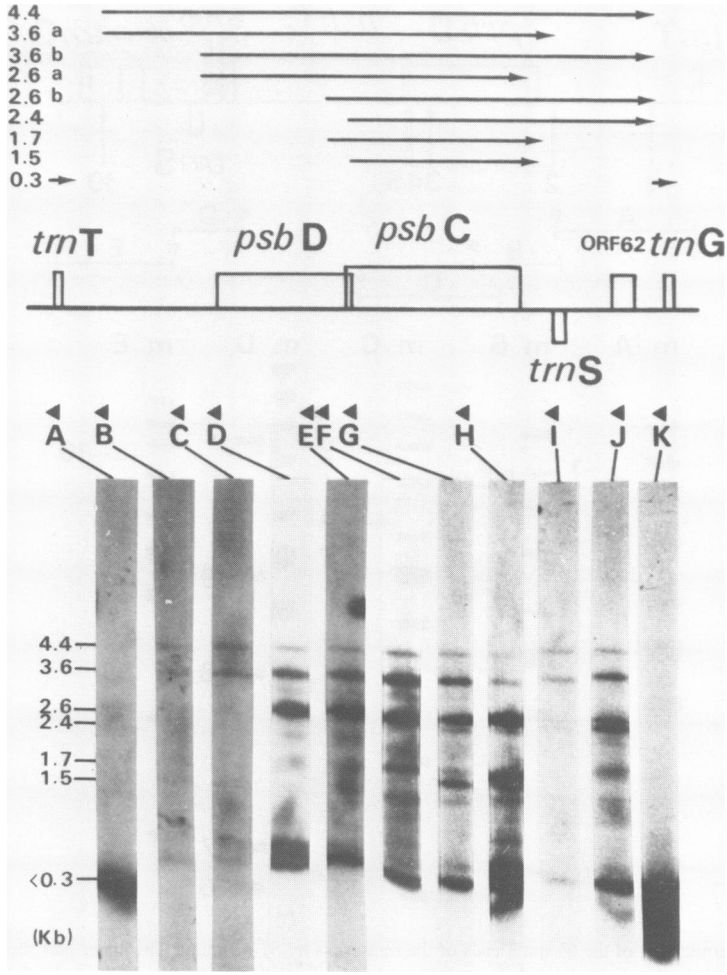

Figure 1. Detection of transcripts from *trn*T, *psb*D, *psb*C, ORF62 and *trn*G. The upper part represents the proposed transcripts. The middle part indicates the gene arrangements and the location of probes used (arrowheads). A, 30 mer (positions 33197–33226); B, 30 mer (positions 33562–33591); C, 31 mer (positions 34171–34201); D, 32 mer (positions 34464–34495); E, 34 mer (positions 35207–35240); F, 30 mer (positions 35325–35354); G, 32 mer (positions 35540–35571); H, 30 mer (positions 36439–36468); I, 30 mer (positions 37174–37203); J, 26 mer (positions 37680–37705); K, 20 mer (positions 38050–38069). The lower part shows autoradiograms of northern blot hybridization. The size marker was a BRL RNA ladder (0.24–9.49 kb).

fragments and degradation products. Probes A and K hybridized only to smaller RNA bands of 0.3 kb or less, indicating that *trn*T and *trn*G are not cotranscribed with the *psb*D-*psb*C cluster.

Determination of the 5' and 3' ends of transcripts

To detect all the 5'ends of transcripts, S1 mapping was first performed using three overlapping DNA fragments [A, B and C] which cover from *trn*T to the middle of *psb*C

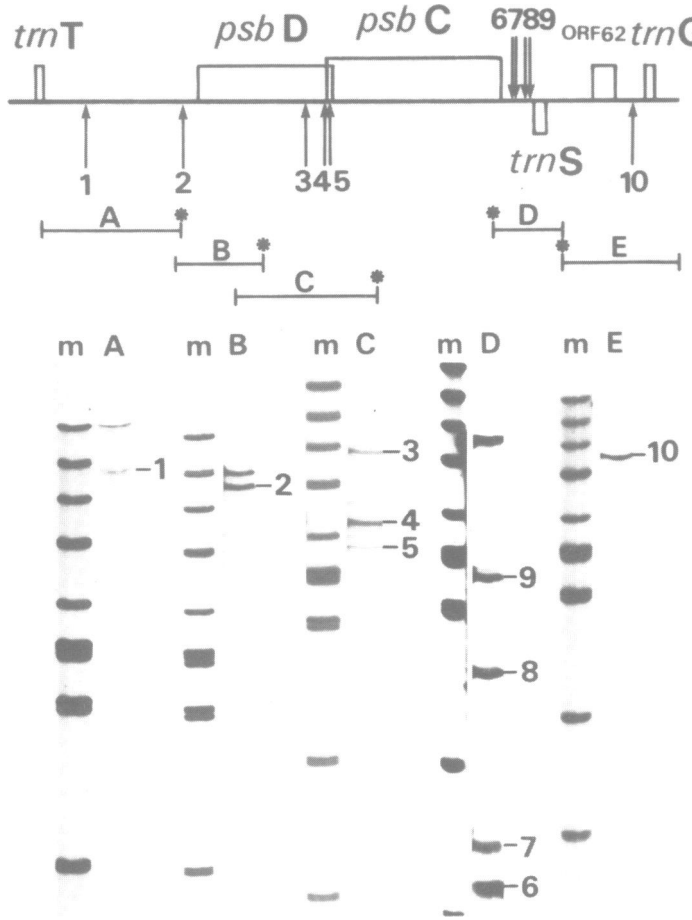

Figure 2. Determination of the 5′ and 3′ends of the transcripts by S1 mapping. The upper part shows the gene arrangement and the location of DNA fragments used (*, ^{32}P-labeled ends). A, 1098 bp *Eco*T14I fragment (positions 33201−34298); B, 705 bp *Dde*I-*Nde*I fragment (positions 34259−35053); C, 1132 bp *Dra*II fragment (positions 34744−35875); D, 555 bp *Ban*II-*Sph*I fragment (positions 36811−37365); E, 941 bp *Sph*I-*Hin*cII fragment (positions 37366−38306). The lower part shows autoradiograms of the protected bands [-**10**] in parallel with a size marker (m, *Hin*cII digest of ϕX174 RF-DNA). The estimated 5′ends [**1**−**5**] and 3′ends [**6**−**10**] are shown above by arrows with the corresponding numerals.

(see Fig. 2). Each of the [5′^{32}P]coding fragments was annealed to total tobacco chloroplast RNA and digested with S1 nuclease. The S1-protected DNA segments were electrophoresed with a size marker. Five 5′ends [**1**−**5**] were detected and their positions on the gene map are shown in Fig. 2.

The precise positions of the 5′ends were then determined by primer extension analysis (Fig. 3). All the 5′ends observed in the S1 mapping analysis were verified by the primer extension analysis and their precise positions on the DNA sequence were at 905 bp and 132 bp upstream from the ATG codon of *psb*D, and at 194 bp upstream, 10 bp upstream and 27 bp downstream from the first ATG codon of *psb*C (Figs. 3 and 5).

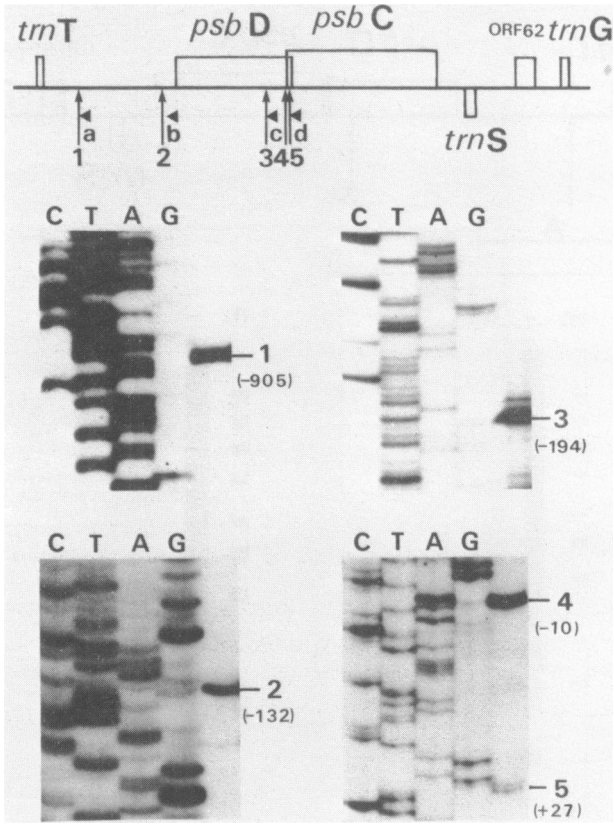

Figure 3. Determination of the 5'ends of the transcripts by primer extension assay. The upper part shows the gene arrangement and the location of primers used (arrowheads). **a**, 32 mer (positions 33661−33692); **b**, 32 mer (positions 34464−34495); **c**, 30 mer (positions 35325−35354); **d**, 30 mer (positions 35581−35610). The 5'ends are indicated above by vertical arrows [**1**−**5**]. The lower part shows autoradiograms of the extended DNAs [**1**−**5**] in parallel with the sequence ladders (C, T, A, G) of the DNA strand complementary to each primer. The positions of the 5'ends relative to the ATG codons of *psb*D and *psb*C are in parentheses.

The 3'ends of transcripts were determined by S1 mapping using the two consecutive fragments [**D** and **E**] labeled at their 3'ends (Fig. 2). The results revealed the presence of four of 3'ends [**6**−**9**] between *psb*C and *trn*S (on the opposite strand) and another 3'end [**10**] between ORF62 and *trn*G, about 1030 bp downstream of the termination codon of *psb*C as shown in Figs. 2 and 5.

Determination of the transcriptional initiation sites.

To determine whether the above 5'ends are transcriptional initiation sites or processed sites, we carried out an additional S1 mapping using *in vitro* capped chloroplast RNA. Total tobacco chloroplast RNA was labeled at the 5'terminal triphosphates with guanylyltransferase in the presence of [α^{32} P]GTP and hybridized to the single stranded DNA (coding strand) derived from subclones in pUC118. The capped RNA/DNA hybrids were digested with single-strand specific P1 nuclease and subjected to size-fractionation in parallel with a size marker.

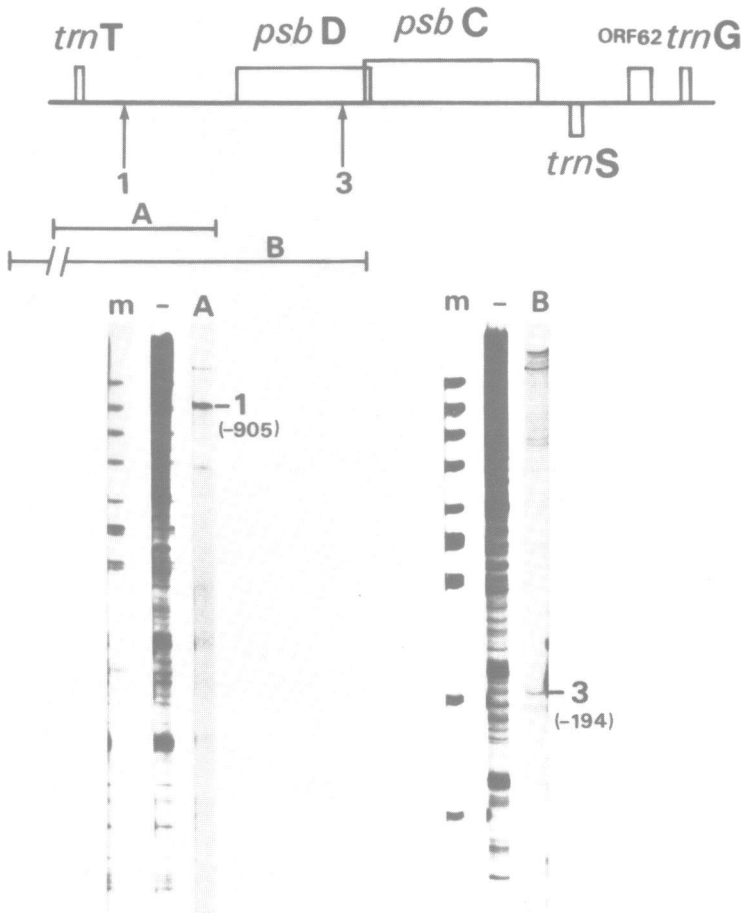

Figure 4. Determination of the transcriptional initiation sites of *psb*D-*psb*C cluster. The upper part shows the gene arrangement and the location of DNA fragments used [**A**, **B**]. The *in vitro* capped RNA was hybridized to the coding DNA strand containing either 1277 bp *Avr*II fragment [**A**, positions 33022−34298] or 3570 bp *Bam*HI-*Eco*81I fragment [**B**, positions 31919−35488]. The heteroduplex was treated with P1 nuclease and electrophoresed on 5% polyacrylamide gels in 8M urea. The lower part shows autoradiograms of the P1 protected RNA fragments [**A**, **B**]. −, capped RNA without P1 treatment; m, size marker (*Hin*cII digest of ϕX174 RF-DNA). The initiation sites [**1**, **3**] are indicated above by vertical arrows.

As shown in Fig. 4, when a single-stranded DNA containing a 1277 bp *Avr*II fragment [**A**, positions 33022−34298] was used, a protected band [**1**] of about 750 nucleotides was obtained and its 5′end corresponds to 905 bp upstream from the *psb*D coding region. We obtained another protected band [**3**] of about 215 nucleotides and its 5′end corresponds to 195 bp upstream of the first ATG codon of *psb*C when another single-stranded DNA containing a 3570 bp *Bam*HI-*Eco*81I fragment [**B**, positions 31918−35488] was used. The latter position is located within the coding region of *psb*D (Fig. 5).

Southern-cross hybridization (two dimensional hybridization) of the *in vitro* capped RNA with DNA fragments containing *psb*D and *psb*C confirmed the above results. Major

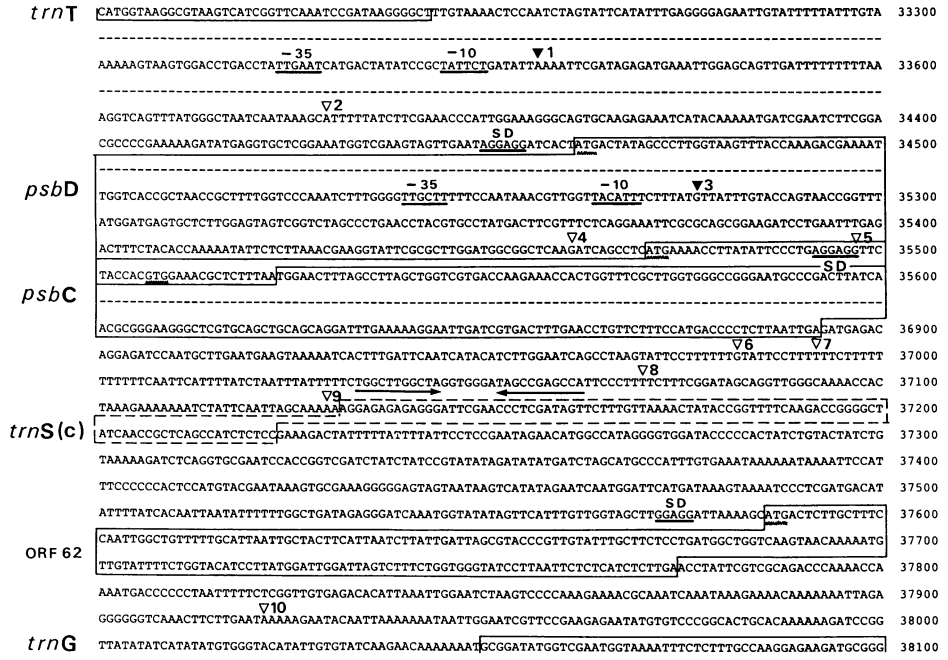

Figure 5. Portions of the nucleotide sequence from *trn*T to *trn*G. Coding regions are boxed. Closed triangles [**1**, **3**], transcriptional initiation sites; open triangles [**2**, **4**, **5**], processing sites; open triangles with bars [**6**–**10**], 3′ends (numbers correspond to those in Figs. 2, 3 and 4). Possible ribosome binding sites (SD), −35 region and −10 regions are underlined. Initiation codons are shown by wavy lines. Inverted repeat sequences are indicated by horizontal arrows. The entire nucleotide sequence was published (24) and deposited with the EMBL database (accession No. Z00044). The deduced amino acid sequences of tobacco *psb*D and *psb*C are presented (25).

transcripts of 4.4, 3.6, 2.6 and 1.7 kb were observed to be hybridized to the *psb*D-*psb*C genes (data not shown). Based on this and the northern blot experiments, we concluded that the 4.4 and 3.6 kb transcripts start from the first initiation site and the 2.6 and 1.7 kb RNAs from the second site.

DISCUSSION

One of the striking features in the plant chloroplast genome is the presence of overlapping genes. The 5′end of *psb*C overlaps the 3′end of *psb*D by 50 nucleotides and the two genes are transcribed in the same direction and translated in different reading frames. Transcription of the *psb*D-*psb*C cluster is highly complex in most higher plants. For example, at least six transcripts were found in spinach (4), and transcription patterns differed between dark-grown and illuminated barley (8). Our northern blot analysis showed the presence of eight major transcripts ranging in size from 1.5 to 4.4 kb; the four transcripts contain both the *psb*D and *psb*C sequences and the other four transcripts contain the *psb*C sequence. The latter transcripts are potentially translatable to produce 43 kDa protein without influencing the *psb*D translation.

RNA heterogeneity could be due to multiple sites of transcriptional initiation. Multiple transcriptional initiation sites were detected for the *cob* and *atp*9 genes in maize mitochondria

by *in vitro* capping experiments. Their positions are in the 5'non-coding region (14). Our *in vitro* capping experiments revealed that there are two transcriptional initiation sites for the *psb*D-*psb*C cluster in tobacco chloroplasts; one [**1**] at 905 bp upstream of *psb*D and the other [**3**] at 194 bp upstream of *psb*C, the letter of which is located within the protein-coding region of *psb*D. The Pribnow box-like sequences and the −35 region-like sequences are found in the regions upstream from the two transcriptional initiation sites (Fig. 5). These sequences are most likely to be transcriptional initiation signals. Internal promoters have been described for several operons in *E. coli* (e.g. 15).

The other three 5'ends [**2**, **4** and **5**], at 132 bp upstream from *psb*D, 10 bp upstream and 27 bp downstream from *psb*C are processing sites. The first processing site [**2**] gives rise to a potential mature mRNA for *psb*D-*psb*C. The second and third processing sites [**4** and **5**] could produce 3'truncated mRNAs for *psb*D, which might direct premature D2-protein molecules. This could result in heterogeneity in length of D2-proteins and explain the diffused nature of D2 proteins in SDS-gel electrophoresis (16) and the discrepancy between molecular weights from ORF (39 kDa) and from electrophoresis (33 kDa). Processing of a mRNA precursor before its termination codon may be an alternative way of peptide chain termination without using termination codons.

The 3'mRNA cleaved off at the second processing site [**4**] is likely to be a mature mRNA for *psb*C. There are two potential initiation codons, ATG and GTG, for *psb*C (Fig. 5). The sequence (10 nucleotides) from the 5'end [**4**] to the first start codon ATG seems too short to bind ribosomes. A sequence AGGAGG that exactly matches the Shine−Dalgarno sequence occurs 9 bp upstream from the second initiation codon GTG while no such sequences before the ATG codon (Fig. 5). These suggest that the GTG functions as the initiation codon for *psb*C rather than the ATG. The tobacco *rps*19 gene is thought to start with GTG (17) and the cyanobacterial *psb*C genes were suggested to use GTG as initiation codons (18, 19).The *psb*C gene from *Chlamydomonas* has recently been sequenced and found to start with GTG and lack the preceeding ATG codon reported in higher plant (20). The N-terminal amino acid sequences of spinach D2 and 43 kDa proteins were determined to start from both N-acetyl-D-phosphothreonines which correspond to the next codons of the ATG (*psb*D) and GTG (*psb*C) (21). This fact supports the GTG being the initiation codon of *psb*C. If the second processing [**4**] gives rise to functional mRNAs for *psb*D and *psb*C, these two genes are not overlapping at the translational level. The third processing [**5**] takes places in the Shine−Dalgarno sequence, suggesting that this cut reduces the synthesis of 43 kDa protein. This may be one of the translational control mechanisms in chloroplasts. A nuclease that cut specifically in the Shine−Dalgarno sequence of some T4 mRNAs has recently been reported (22).

Our S1 mapping revealed that there are five 3'ends for the transcripts; between *psb*C and *trn*S, and between ORF62 and *trn*G (Fig. 5). There is an inverted repeat, which can form a stable hairpin, between *psb*C and ORF62. Stern and Gruissem reported that inverted repeat sequences found at the 3'ends of spinach plastid transcription units can serve as efficient RNA processing elements (23). If this is the case, one or more of the 3'ends mapped bet veen *psb*C and *trn*S are processing sites.

The largest transcript detected by our Southern-cross hybridization and northern blot experiment is 4.4 kb. The length from the transcriptional initiation site (905 bp upstream of *psb*D) to the last 3'end (1030 bp downstream of *psb*C) is 4.36 kb. No large transcripts were detected by probe K corresponding to *trn*G (Fig. 1) and no 3'ends were observed in the region downstream of *trn*G (Fig. 2). Based on these observations the 3'end between

ORF62 and *trn*G is most likely to be the transcriptional termination site. The tobacco *psb*D-*psb*C cluster includes ORF62 as a transcription unit but not *trn*G.

ACKNOWLEDGEMENTS

We thank T. Wakasugi, N. Hayashida, T. Matsubayashi and K. Torazawa for their help and suggestions. This work was principally supported by a Grant-in-Aid from the Ministry of Education, Science and Culture (Japan). B.Y. Meng is supported by Makita Scholarship Foundation.

REFERENCES

1. Herrmann, R.G., Westhoff, P., Alt, J., Tittgen, J. and Nelson, N. (1985) in Molecular Form and Function of the Plant Genome, ed. by L. van Vloten-Doting, G.S.P. Groot and T.C. Hall, pp.233–256. Plenum, New York.
2. Gray, J.C. (1987) in Photosynthesis, ed. by J. Amesz, pp.319–342. Elsevier, Amsterdam.
3. Sugiura, M. (1989) Ann. Rev. Cell Biol. 5, 51–70.
4. Alt, J., Morris, J., Westhoff P. and Herrmann, R.G. (1984) Curr. Genet. 8, 507–606.
5. Holschuh, K., Bottomley, W. and Whitfeld, O.R. (1984) Nucl. Acids Res. 12, 8819–8834.
6. Rasmussen, O.F., Bookjans, G., Stummann, B.M. and Henningsen, K.W. (1984) Plant Mol. Biol. 14, 395–400.
7. Berends, T., Gamble, P.E. and Mullet, J.E. (1987) Nucl. Acids. Res. 15, 5217–5240.
8. Gamble, P.G., Berends, T. and Mullet, J.E. (1988) EMBO J. 7, 1289–1297.
9. Woodbury, N.W., Roberts, L.L., Palmer, J.D. and Thompson, W.F. (1988) Curr. Genet. 14, 75–89.
10. Sugiura, M., Shinozaki, K., Zaita, N., Kusuda, M. and Kumano, M. (1986) Plant Sci. 44, 211–216.
11. Tanaka, M., Obokata, J., Chunwongse, J., Shinozaki, K. and Sugiura, M (1987) Mol. Gen. Genet. 29, 427–431.
12. Shinozaki, K. and Sugiura, M. (1982) Gene 20, 91–102.
13. Meng, B.Y., Tanaka, M., Wakasugi, T., Ohme, M., Shinozaki, K. and Sugiura, M. (1988) Curr. Genet. 14, 395–400.
14. Mulligan, R.M., Maloney, A.P. and Walbot, V. (1988) Mol. Gen. Genet. 211, 373–380.
15. Lopes, J.M. and Lawther, R.P. (1986) Nucl. Acids Res. 14, 2779–2798.
16. Chua, N.H. and Gillham, N.W. (1977) J. Cell Biol. 77, 441–452.
17. Sugita, M. and Sugiura, M. (1983) Nucl. Acids. Res. 6, 1913–1918.
18. Chisholm, D. and Williams, J.G.K. (1988) Plant Mol. Biol. 10, 293–301.
19. Golden, S.S. and Stearns, G.W. (1988) Gene 67, 85–96.
20. Rochaix, J.D., Kuchka, M., Mayfield, S., Schirmer-Rahire, M., Girard-Bascou, J. and Bennoun, P. (1989) EMBO J. 8, 1013–1021.
21. Michel, H., Hunt, D.F., Shabanowitz, J. and Bennett, J. (1988) J. Biol. Chem. 263, 1123–1130.
22. Uzan, M., Favre, R. and Brody, E. (1988) Proc. Natl. Acad. Sci. USA 85, 8895–8899.
23. Stern, D.B. and Gruissem, M. (1987) Cell 51, 1145–1157.
24. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
25. Shinozaki, K., Hayashida, N. and Sugiura, M. (1988) Photosynthesis Res. 18, 7–31.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.