Curr Genet (1991) 20:259–264

Current Genetics

Two promoters within the *psbK-psbI-trnG* gene cluster in tobacco chloroplast DNA

Bing-Yuan Meng, Tatsuya Wakasugi, and Masahiro Sugiura

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

Received April 29, 1991

Summary. Transcription of the 2.6 kbp *psbK-psbI-trnG* cluster in tobacco chloroplasts has been studied. This cluster contains, in linear sequence, the genes encoding two low-molecular-mass polypeptides, K and I, of photosystem II (*psbK* and *psbI*, respectively), and $tRNA^{Gly}$ (UCC) (*trnG*). Northern blot hybridization revealed that the largest transcript (2.6 kb) hybridizes to *psbK*, *psbI* and *trnG*, but not to the following *trnR*-UCU. Ten other transcripts ranging from 0.1 to 1.3 kb were also detected. Three of these transcripts overlap the divergent transcript arising from *trnS*-GCU located on the opposite DNA strand. S1 mapping and primer extension experiments showed that these multiple transcripts comprise eight distinct 5′ ends. By in vitro capping assays two of them were determined to be transcriptional initiation sites; one is located 163 bp upstream of *psbK* and the other is 6 bp upstream of *trnG*. The 3′ ends of transcripts were determined by S1 mapping; one lies between *psbI* and *trnG* and the other is at the end of *trnG*. The presence of dual promoters of *trnG* is discussed.

Key words: Transcription – Promoter – *psbK* operon – In vitro capping – Tobacco

Introduction

Photosystem II (PSII) is one of the four supramolecular multiprotein complexes of chloroplast thylakoid membranes which are involved in the light reaction of photosynthesis (Herrmann et al. 1985; Gray 1987). PSII contains about 20 different polypeptides of which at least 12 are encoded in the chloroplast genome (for review see Sugiura 1989). These include: D1 (*psbA*) and D2 (*psbD*) polypeptides, which bind the chlorophyll, pheophytin and plastoquinone molecules of the reaction center; 47 kDa (*psbB*) and 43 kDa (*psbC*) chlorophyll-apoproteins; 10 kDa phosphoprotein (*psbH*); two subunits of cytochrome b559 (*psbE*, *psbF*) and five low-molecular mass PSII polypeptides (*psbI*, *K*, *L*, *M* and *N*).

Two ORFs of 98 and 52 codons have been found in the opposite strand between *trnS*-GCU and *trnQ*-UUG in the tobacco chloroplast genome (Deno and Sugiura 1983). ORF98 was identified as the gene for a PSII component (*psbK*) by comparison with the partial amino acid sequence of the 2 kDa polypeptide of spinach (Murata et al. 1988). ORF98 contains two potential start codons, GTG at position 1 and ATG at position 38. Comparison with the corresponding ORFs from several other plant suggests that the ATG is likely to be the initiation codon and, hence, tobacco *psbK* probably consists of 61 codons (ORF61). ORF52 was also identified as the gene for another PSII component (*psbI*) by determination of the partial amino acid sequence of the spinach 4.8 kDa polypeptide (Ikeuchi and Inoue 1988) and the pea 4.5 kDa polypeptide (Webber et al. 1989). Downstream from *psbK-psbI* is *trnG*-UCC (containing a 691 bp intron in the D-stem region) and then *trnR*-UCU in the same direction (Deno and Sugiura 1984). Therefore, the tobacco (dicot) gene organization is (*trnQ*)-*psbK-psbI*-(*trnS*)-*trnG-trnR* (parenthesized genes on the opposite DNA strand, see Fig. 1, Shinozaki et al. 1986). However, the gene organization in monocots (wheat, rice and barley), (*trnQ*)-*psbK-psbI*-(*trnS*)-*psbD-psbC*-(*trnS*-UGA) differs (Howe et al. 1988; Hiratsuka et al. 1989; Sexton et al. 1990a). This can be explained by the rearrangement of the monocot LSC with respect to the dicot LSC which results from at least three inversions in a tobacco-like ancestral form (for review see Sugiura 1989).

As *psbK-psbI* was reported to be co-transcribed with the downstream *psbD-psbC* in barley (Sexton et al. 1990a), it is of interest to know whether this is also the case for the dicot genome. We have analyzed transcripts from this region of tobacco and detected a large transcript of 2.6 kb covering *psbK-psbI-trnG* but not *trnR*. Using an in vitro capping method, we found two initiation sites in this gene cluster: one is located upstream of *psbK* and the other is immediately upstream of *trnG*.

Offprint requests to: M. Sugiura

Materials and methods

Recombinant plasmid pTB24, containing the 13.2 kbp *Bam*HI partial fragment (Ba26c-Ba10a-Ba4), has been described previously (Sugiura et al. 1986). Appropriate short DNA pieces derived from the plasmid DNA were either subcloned into pUC118/119 vectors or extracted from gels after electrophoresis. Oligonucleotides of 23–36 mers were prepared by a DNA synthesizer (Applied Biosystems, USA, 380A) and labeled at their 5′ ends with ^{32}P. Tobacco (*Nicotiana tabacum* var. Bright Yellow 4) was grown in a greenhouse for 2–3 months and total chloroplast RNA was prepared as described by Tanaka et al. (1987). For Northern blot hybridization, it was separated in 1.1% agarose/6% formaldehyde gels, blotted onto a nylon membrane sheet (Amersham, UK, Hybond N) and subsequently hybridized with [5′ ^{32}P]oligonucleotide probes according to BioRad's, USA instruction manual except that our prehybridization solution contained 1% SDS instead of 7%. S1 mapping was performed as described (Shinozaki and Sugiura 1982). 3′ end-labeling was carried out using a DNA 3′ end-labeling kit (Boehringer, Mannheim, FRG) and [α ^{32}P]ddATP (185 TBq/mmol). Primer extension was carried out using Moloney murine leukemia virus reverse transcriptase (BRL, USA) as described by Meng et al. (1988). For in vitro capping, the 5′ terminal triphosphates of total tobacco chloroplast RNA were labeled with [α ^{32}P]GTP (110 TBq/mmol) and vaccinia virus guanylyltransferase (BRL) as previously described (Kennell and Pring 1989).

Results

Detection of transcripts

To detect transcripts from *psbK* (61 codons), *psbI* (52 codons), *trnG* and *trnR*, Northern blot analysis was performed using total tobacco chloroplast RNA and [5′ ^{32}P]oligonucleotide probes (probes A–I). As shown in Fig. 1, probes B–H hybridized to a common RNA band of 2.6 kb. This band was not detected with probe A (in the spacer between *trnQ* and *psbK*) and probe I (*trnR*), indicating that the 2.6 kb transcript starts after *trnQ* and ends before *trnR*. Probes B–H, which span 2.5 kbp, hybridized also to a 1.3 kb RNA band, suggesting that there are two or more RNA species which have a similar size of 1.3 kb but differ in their ends (designated as 1.3 a and 1.3 b in Fig. 1). Probes B–D hybridized to a 0.7 kb RNA and probes B and C hybridized to a 0.4 kb RNA, both of which encode the K-peptide. A 0.9 kb RNA band was detected with probes D–E but not with probes A–C, suggesting that it has a 5′ end after *psbK* and encodes the I-peptide. RNA bands of 1.3 b, 1.15, 0.85, and an 0.75 kb RNA band, were found with probes F–H, and these RNAs are likely to be the precursors of $tRNA^{Gly}$ (UCC), which contains a 691 bp intron in the D-stem region (Deno and Sugiura 1984). Another 0.7 kb RNA band was detected with probes G and H but not with probe F, and this RNA may be a splicing intermediate containing the intron and the 3′ exon. Two minor bands of 2.0 and 1.5 kb were detected together with a mature $tRNA^{Gly}$ band of 0.1 kb with probes F and H, but not with probes E, G and I. The former two bands may be the result of cross-hybridization to the precusors of other $tRNAs^{Gly}$ which have significant homology with probes F and H. Probe I hybridized only to a dense RNA band of about 0.1 kb, indicating that *trnR* is not co-transcribed with the *psbk-psbI-trnG* cluster.

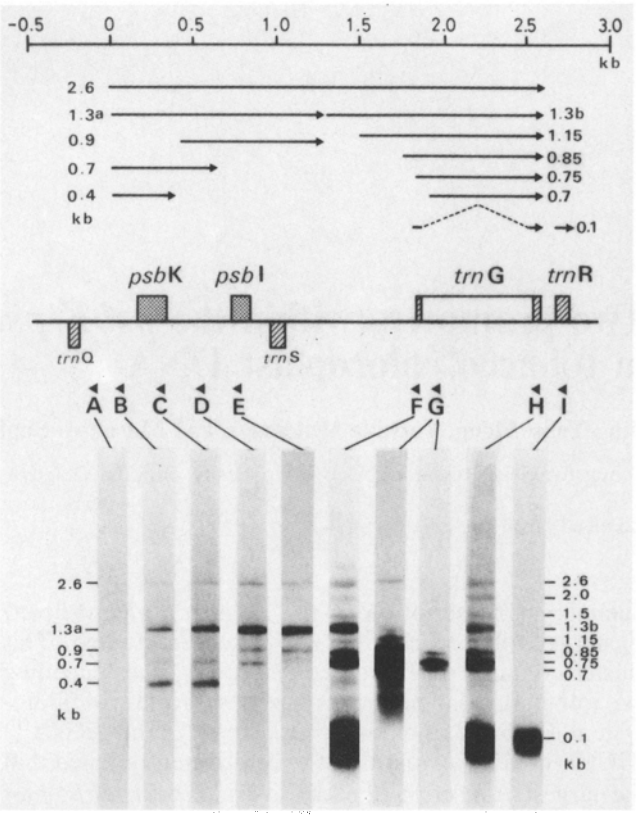


Fig. 1. Detection of transcripts from *psbK*, *psbI*, *trnG* and *trnR*. The upper part represents the proposed transcripts. The middle part indicates the gene arrangements and the location of probes used (*arrowheads*). *A*, 36-mer (positions -277 to -241); *B*, 31-mer (positions -112 to -81); *C*, 29-mer (positions 64–93); *D*, 29-mer (positions 363–392); *E*, 30-mer (positions 564–594); *F*, 23-mer (positions 1663–1686); *G*, 33-mer (positions 1785–1818); *H*, 26-mer (positions 2391–2417); *I*, 27-mer (positions 2630–2657). The lower part shows autoradiograms of Northern blot hybridization. Two autoradiograms with probe G are shown; long exposure (left) and short exposure (right). The size marker was a BRL RNA ladder (0.24–9.49 kb)

Transcripts from *trnS* and *trnQ*, which are located on opposite DNA strands, were also detected by Northern blot analysis (unpublished data), indicating that a region containing *trnS* is divergently transcribed.

Determination of the 5′ and 3′ ends of transcripts

Primer extension analysis was conducted to determine the 5′ ends of the above transcripts. Oligonucleotides B, D, K, G and I were used as primers, the sites of which are internal to the major transcripts (Fig. 2). Total tobacco chloroplast RNA was hybridized to each of the primers labeled with ^{32}P at their 5′ ends, extended by reverse transcription and then digested with RNase A. The extended DNA was electrophoresed in parallel with the sequence ladder obtained by the dideoxy method using the same primer. Eight extended bands were detected and their positions are shown on the gene map (Fig. 2). The precise positions of these bands are at -163 bp upstream of the ATG codon of *psbK* (1), -315 bp from *psbI* (2),

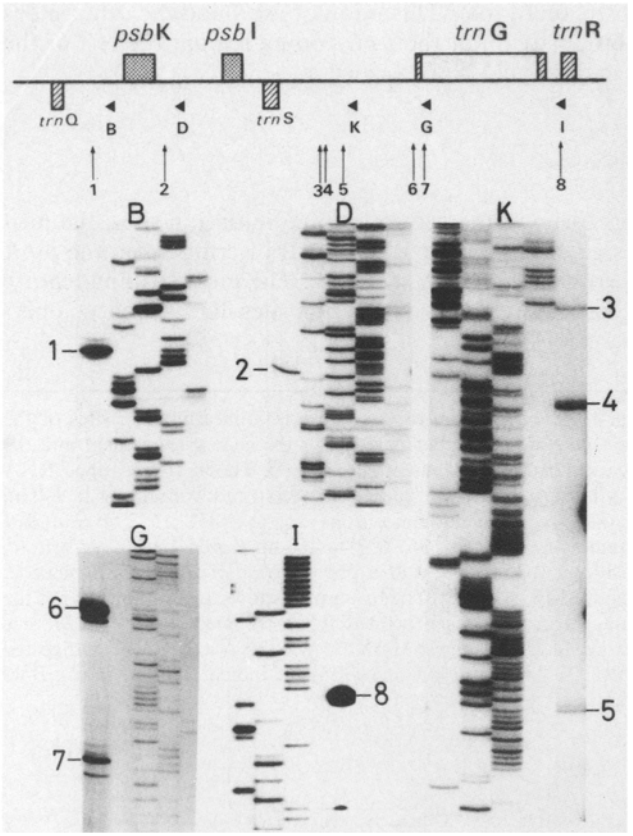


Fig. 2. Determination of the 5′ ends of transcripts by primer extension assay. The upper part shows the gene arrangement and the location of primers used (*arrowheads*). The position of probes *B, D, G, I* are the same as Fig. 1. Probe *K*, 27-mer (positions 1370–1397). The 5′ ends are indicated by *vertical arrows* [1–8]. The lower part shows autoradiograms of the extended DNAs [1–8] in parallel with the sequence ladders of the DNA strand complementary to each primer

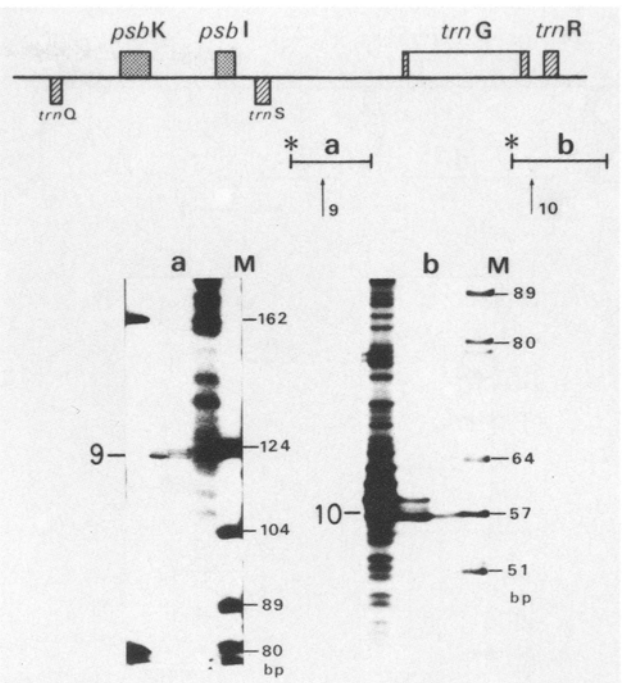


Fig. 3. Determination of the 3′ ends of transcripts by S1 mapping. The upper part shows the gene arrangement and the location of DNA fragments used (*, ^{32}P-labeled 3′ ends). *a*, 467 bp *Ban*II/*Dra*I fragment (positions 1031–1498); *b*, 559 bp *Sal*I/*Sna*BI fragment (positions 2370–2929). The lower part shows autoradiograms of the protected bands (*9, 10*) in parallel with a size marker (*M*, a *Hae*III digest of pBR322). The 3′ ends (*9, 10*) are shown by *vertical arrows*

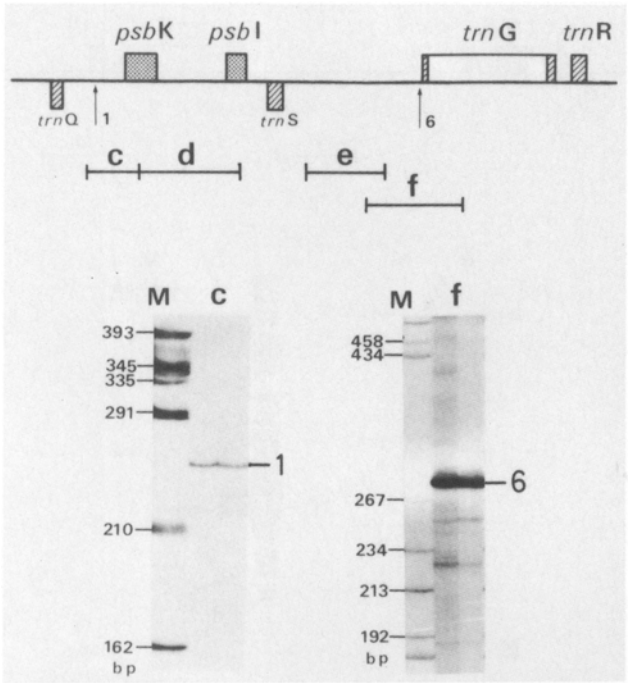
-510, -465, -378, -6 and +24 bp from the 5′ exon of *trnG* (3, 4, 5, 6 and 7, respectively), and +1 bp from *trnR* (8) as shown in Figs. 2 and 5. These termini correlate with those of the major transcripts approximated by Northern hybridization analysis. All the 5′ ends observed in the primer extension analysis were verified by S1 mapping (data not shown).

The 3′ ends of transcripts were determined by S1 mapping using total tobacco chloroplast RNA and two DNA fragments, **a** (*Ban*II/*Dra*I, 467 bp) and **b** (*Sal*I/*Sna*BI, 559 bp), labeled with ^{32}P at their 3′ ends (Fig. 3). The results revealed the presence of two 3′ ends: one between *psbI* and *trnG* and the other at the end of *trnG*. Their positions on the DNA sequence are 483 bp downstream from the termination codon of *psbI* and 2 bp downstream from the 3′ exon of *trnG* as shown in Figs. 3 and 5. Based on the results of our Northern blot hybridization (Fig. 1) the other two 3′ ends (of the 0.7 and 0.4 kb RNAs) are expected to be located between *psbK* and *psbI*. We failed to detect clear S1 protected bands for the 3′ ends of the 0.7 and 0.4 kb RNAs, probably because this region is rich in AT.

Determination of the transcriptional initiation sites

Chloroplast primary transcripts retain triphosphates at their 5′ ends which can be specifically capped in vitro. 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 RNAs. Total tobacco chloroplast RNA was labeled at the 5′ terminal triphosphates with guanylyltransferase in the presence of [α ^{32}P]GTP and then hybridized to each of the DNA fragments, **c**, **d**, **e** and **f** (Fig. 4). The capped RNA/DNA hybrids were treated with single-strand specific mung bean nuclease and subjected to size-fractionation in parallel with a size marker. When a 300 bp *Eco*RI/*Stu*I fragment (**c**, positions -215 to +85) was used, a protected RNA band of about 245 nucleotides was obtained and its 5′ end corresponds to 163 bp upstream of the *psbK* coding region. Another protected band of about 215 nucleotides was observed by using a 560 bp *Hin*dIII/*Bam*HI fragment (**f**, positions 1375–1935) and its 5′ end corresponds to 6 bp upstream of the 5′ exon of *trnG*. No protected RNA bands were obtained when 572 bp *Stu*I/*Bst*NI (**d**, positions 85–657) and 463 bp *Bam*HI/*Dra*I (**e**, positions 1 036–1 499) fragments were used (data not shown).

We concluded, therefore, that there are two initiation sites in the *psbK-psbI-trnG* cluster: one is located 163 bp



upstream of the ATG codon of *psbK* and the other site is 6 bp upstream of the *trnG* coding region; the rest of the 5′ ends should be processed sites.

Discussion

Fig. 5 shows the transcriptional initiation sites, the processed sites and the 3′ ends of transcripts from the *psbK* operon in the DNA sequence. The most striking feature is the presence of two initiation sites in this tobacco oper-

Fig. 4. Determination of the transcriptional initiation sites of the *psbK* cluster. The upper part shows the gene arrangement and the location of DNA fragments used (c–f). The in vitro capped RNA was hybridized to the coding DNA strand containing a 300 bp *Eco*RI/*Stu*I fragment (*c*, positions -215 to +85), a 572 bp *Stu*I/*Bst*I fragment (*d*, positions 85–657), a 463 bp *Bam*HI/*Dra*I fragment (*e*, positions 1036–1499) and a 560 bp *Hin*dIII/*Bam*HI fragment (*f*, positions 1375–1935). The lower part shows autoradiograms of the mung bean nuclease protected RNA fragments *1* and *6*. *M*, size marker (a *Hin*cII digest of øX174 RF-DNA and a *Hae*III-digested pBR322). The initiation sites (*1, 6*) are indicated by *vertical arrows*

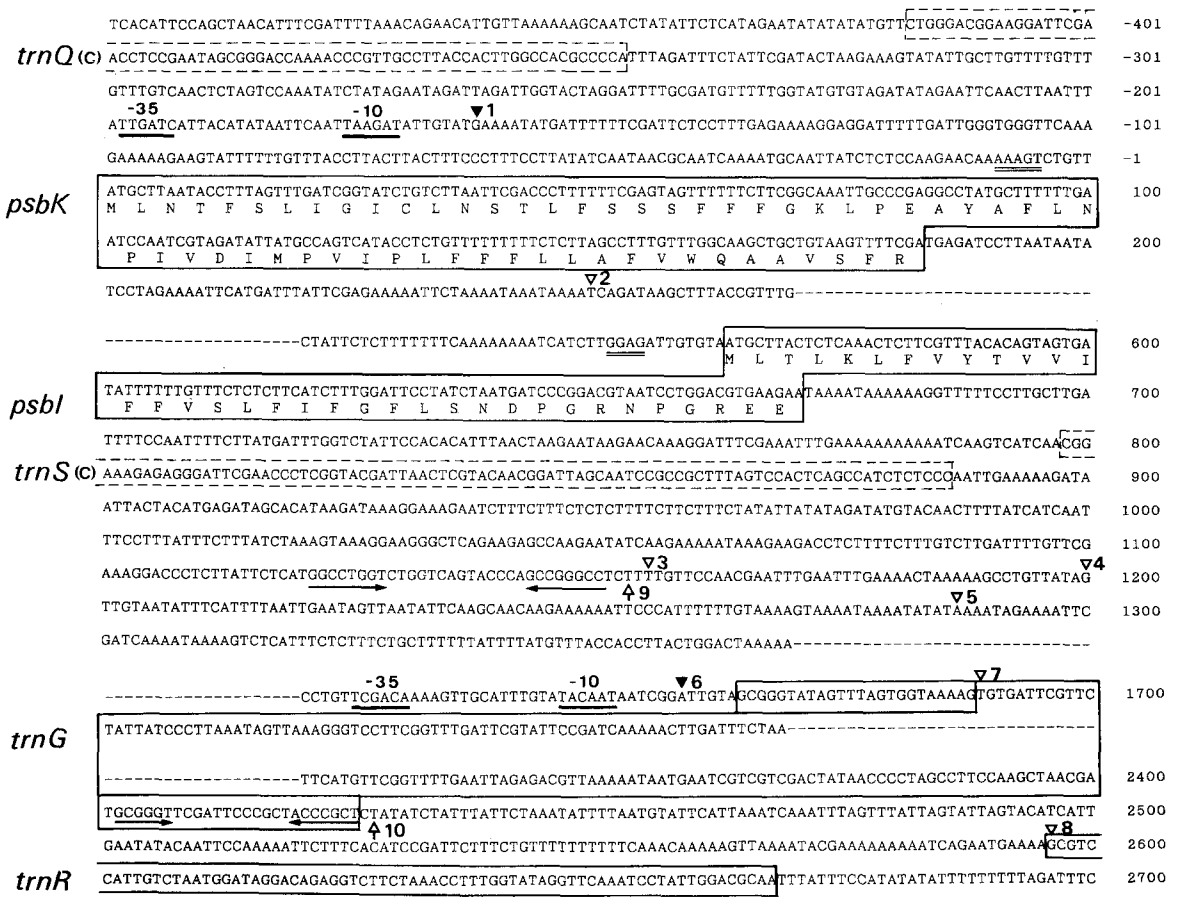


Fig. 5. Portions of the nucleotide sequence from *psbK* to *trnR*. Coding regions are *boxed*. *Closed triangles* (*1, 6*), transcriptional initiation sites; *open triangles* (*2, 3, 4, 5, 7*), processed sites; *open triangles with bars* (*9, 10*), 3′ ends (numbers correspond to those in Figs. 2, 3 and 4). Possible ribosome binding sites (SD), -35 regions and -10 regions are *underlined*. Inverted repeat sequences are indicated by *horizontal arrows*. The entire nucleotide sequence has been published (Shinozaki et al. 1986) and deposited with the EMBL database (accession No. Z00044)

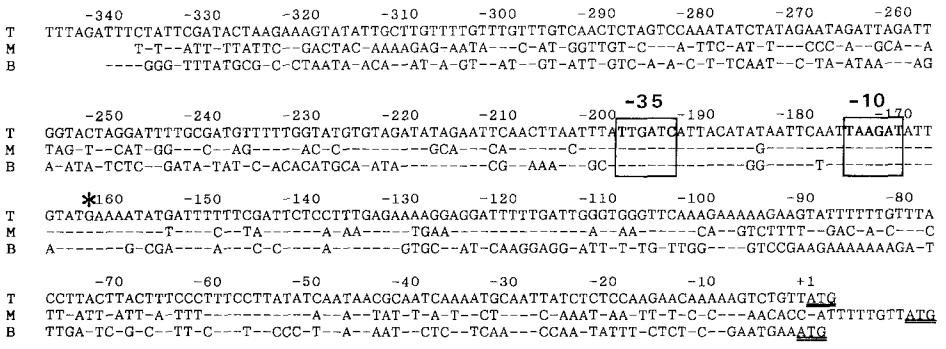


Fig. 6. Comparison of the upstream sequences of the *psbK* operon from tobacco (*T*), mustard (*M*) and barley (*B*). The nucleotide sequences between *trnQ* and *psbK* are shown (identical nucleotides are *hyphoned*). The transcriptional initiation site of tobacco is indicated with an *asterisk*. The -35 and -10 elements are *boxed*. Initiation codons are *underlined*

on, one in front of *psbK* and the other immediately before *trnG*. In barley, *psbK* has been shown to be co-transcribed with *psbI*, *psbD* and *psbC*, and the 5′ end of the transcript is about 150 bp upstream of *psbK* (Sexton et al. 1990a). In mustard, the transcriptional start site has been reported to be about 520 bp upstream of *psbK* and likely to overlap that of *rps16* on the opposite strand (Neuhaus and Link 1990). These sites have been mapped by either primer extension or S1 mapping, neither of which can discriminate initiation sites from processed sites. We have demonstrated the transcriptional initiation sites of the tobacco *psbK* operon by in vitro capping experiments which only detect initiation sites. The first position (-163) is similar to the 5′ end of the corresponding transcript in barley, but not in mustard. The Pribnow box-like sequence (TAAGAT) and the -35 region-like sequence (TTGATC) are found upstream of the first initiation site. These motifs and their surrounding sequences are highly conserved among tobacco, barley and mustard (Fig. 6). It is, therefore, peculiar that mustard does not have the initiation site in this region. An additional cis-element may be necessary to initiate transcription in mustard.

Our Northern blot hybridization showed that the largest transcript of 2.6 kb covers *psbK*, *psbI* and *trnG* but not *trnR*, suggesting that the expression of *trnG* is co-regulated with *psbK-psbI*. Further analysis, however, revealed that there is an additional promoter in front of *trnG*. This initiation site is 6 bp upstream of *trnG* and is identical with the 5′ end of a transcript from *trnG* determined previously by S1 mapping (Deno and Sugiura 1984). The steady state level of *trnG* transcripts is much higher than those of *psbK-psbI*, suggesting that transcriptional efficiency differs significantly between *trnG* and *psbK-psbI* promoters. In barley, the largest transcript of *psbK-psbI* covers the downstream *psbD-psbC* and ORF62 (Sexton et al. 1990a) although *psbD-psbC* has its own promoter (Yao et al. 1989; Sexton et al. 1990b). A light-induced switch from the *psbK* promoter to the *psbD-psbC* promoter has been proposed in barley (Sexton et al. 1990b). As in the case of barley *psbD-psbC*, tobacco *trnG* is transcribed from two distinct promoters. The utilization of the dual promoters may be differentially regulated. Another explanation for the dual promoters is that there is no strong termination signal downstream from *psbI* so that transcription proceeds to the next gene. Transcripts from the first and second initiation sites most likely use the same termination signal and our S1 mapping assay revealed that it is the 3′ end of *trnG*. This may also be the case for barley. Transcripts from three different initiation sites in barley may use one termination signal (Sexton et al. 1990a).

Our S1 mapping revealed that there is another 3′ end between *psbI* and *trnG* (9 in Fig. 3). The region immediately preceding this site can potentially form a stem-and-loop secondary structure (Fig. 5). Stem-and-loop structures have been shown to be ineffective as transcription terminators, but serve as efficient RNA processing elements (Stern and Gruissem 1987). Therefore, this site is probably a processing site.

Acknowledgements. We thank Drs. M. Sugita, M. Tanaka and T. Hidaka for suggestions and help. B-Y. Meng was funded by a Monbusho fellowship.

References

Deno H, Sugiura M (1983) Nucleic Acids Res 11:8407–8414
Deno H, Sugiura M (1984) Proc Natl Acad Sci USA 81:405–408
Gray JC (1987) In: Amesz J (ed) Photosynthesis. Elsevier, Amsterdam, pp 319–342
Herrmann RG, Westhoff P, Alt J, Tittgen J, Nelson N (1985) In: Van Vloten-Doting L, Groot GSP, Hall T (eds) Molecular form and function of the plant genome. Plenum, New York, pp 233–256
Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li Y, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) Mol Gen Genet 217:185–194
Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) Curr Genet 13:343–349
Ikeuchi M, Inoue Y (1988) FEBS Lett 241:99–104
Kennell JC, Pring DR (1989) Mol Gen Genet 216:16–24
Meng BY, Tanaka M, Wakasugi T, Ohme M, Shinozaki K, Sugiura M (1988) Curr Genet 14:395–400
Murata N, Miyao M, Hayashida N, Hidaka T, Sugiura M (1988) FEBS Lett 235:283–288
Neuhaus H, Link G (1990) Curr Genet 18:377–383
Sexton TB, Jones JT, Mullet JE (1990a) Curr Genet 17:445–454
Sexton TB, Christopher DA, Mullet JE (1990b) EMBO J 9:4485–4494
Shinozaki K, Sugiura M (1982) Gene 20:91–102
Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Shinozaki-

Yamaguchi 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 (1986) EMBO J 5:2043–2049
Stern DB, Gruissem W (1987) Cell 51: 1145–1157
Sugiura M (1989) Annu Rev Cell Biol 5:51–70
Sugiura M, Shinozaki K, Zaita N, Kusuda M, Kumano M (1986) Plant Sci 44:211–216
Tanaka M, Obokata J, Chunwongse J, Shinozaki K, Sugiura M (1987) Mol Gen Genet 209:427–431
Webber AN, Packman L, Chapman DJ, Barber J, Gray JC (1989) FEBS Lett 242:259–262
Yao WB, Meng BY, Tanaka M, Sugiura M (1989) Nucleic Acids Res 17:9583–9591

Communicated by H. Kössel