Curr Genet (1990) 17:445–454

Current Genetics

Sequence and transcriptional analysis of the barley ctDNA region upstream of *psb*D-*psb*C encoding *trn*K(UUU), *rps*16, *trn*Q(UUG), *psb*K, *psb*I, and *trn*S(GCU)

Tineke Berends Sexton[1], Jennifer T. Jones[2], and John E. Mullet[2]

[1] M. D. Anderson Hospital, Department of Tumor Biology, Houston, TX 77030, USA
[2] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA

Received December 1, 1989

Summary. A 6.25 kbp barley plastid DNA region located between *psb*A and *psb*D-*psb*C was sequenced and RNAs produced from this DNA were analyzed. *Trn*K(UUU), *rps*16 and *trn*Q(UUG) were located upstream of *psb*A. These genes were transcribed from the same DNA strand as *psb*A and multiple RNAs hybridized to them. *Trn*K and *rsp*16 contained introns; a 504 amino acid open reading frame (ORF504) was located within the *trn*K intron. Between *trn*Q and *psb*D-*psb*C was a 2.24 kbp region encoding *psb*K, *psb*I and *trn*S(GCU). *Psb*K and *psb*I are encoded on the same DNA strand as *psb*D-*psb*C whereas *trn*S(GCU) is transcribed from the opposite strand. Two large RNAs accumulate in barley etioplasts which contain *psb*K, *psb*I, anti-sense *trn*S(GCU) and *psb*D-*psb*C sequences. Other RNAs encode *psb*K and *psb*I only, or *psb*K only. The divergent *trn*S(GCU) located upstream of *psb*D-*psb*C and a second divergent *trn*S(UGA) located downstream of *psb*D-*psb*C were both expressed. Furthermore, RNA complementary to *psb*K and *psb*I mRNA was detected, suggesting that transcription from divergent overlapping transcription units may modulate expression from this DNA region.

Key words: Barley – Plastid, DNA, RNA

Introduction

Higher plant plastids contain multiple copies of a circular genome (Bendich 1987). Several genomes have now been sequenced and were found to encode 120–140 genes (Shinozaki et al. 1986b; Ohyama et al. 1986; Hiratsuka et al. 1989). Many of the plastid genes encode proteins or RNAs involved in transcription or translation. Among this group of genes are *rpo*A, *rpo*B, *rpo*C_1, and *rpo*C_2, which encode RNA polymerase subunits, 18 genes encoding ribosomal proteins, 30 tRNA genes and 4 rRNA-encoding genes. A second large group of plastid genes encodes proteins involved in photosynthesis. The majority of this group encode thylakoid membrane proteins which are associated with the photosynthetic electron transport complexes; Photosystem I (PSI), Photosystem II (PSII), a cytochrome b_6/f complex and an ATP synthetase. Other proteins found in these complexes are encoded by nuclear genes. Therefore, coordinated expression of plastid genes and nuclear genes encoding plastid proteins is required for efficient biosynthesis of the photosynthetic electron complexes.

Numerous proteins of PSII are encoded by plastid genes. These include D_1 and D_2, which bind chlorophyll, pheophytin and quinone, and form a heterodimer which mediates primary charge separation and additional electron transfer reactions. D_1 and D_2 are encoded by *psb*A and *psb*D respectively. In barley, these genes are located approximately 8 kbp apart near one inverted repeat (Berends et al. 1987). A third PSII gene, *psb*C, encodes a 43 kD chlorophyll-apoprotein. *Psb*C is located adjacent to *psb*D and the *psb*C open reading frame may overlap that of *psb*D by 50 bp (Neumann 1988; Efimov et al. 1988a). The tandem arrangement of *psb*D-*psb*C is found in plastids of all vascular plants examined to date. Other plastid genes which encode PSII proteins include *psb*B (encodes a 47 kD chlorophyll-apoprotein), *psb*H (encodes a 10 kD phosphoprotein), *psb*E-*psb*F (encode subunits of cyt559), and *psb*I through *psb*N which encode low molecular weight PSII proteins (< 10 kD) (Alt et al. 1983; Westhoff et al. 1986; Webber et al. 1989a; Murata et al. 1988; Webber et al. 1989b; Moller et al. 1989).

The expression of several genes encoding PSII proteins has been investigated. Early studies focused on *psb*A because RNA produced from this gene accumulates to high levels in illuminated plants (Bedbrook et al. 1978; Link 1982). In barley, *psb*A transcription was found to initiate 80 bp upstream from the coding region. In vitro transcription analysis revealed that the *psb*A promoter contained DNA sequences –10 and –35 from the transcription initiation site which were homologous to bacterial promoter elements (Boyer and Mullet 1988). The

Offprint requests to: J. E. Mullet

transcription unit which encodes *psb*B and *psb*H has also been examined in some detail. These genes are cotranscribed with *pet*D and *pet*B, genes encoding subunits of the cytochrome b_6/f complex. A large number of RNAs are produced from this transcription unit. This RNA heterogeneity arises from two RNA splicing events, processing of the 5′ untranslated RNA preceding the *psb*B coding region and several presumptive internal cleavage sites (Westhoff 1985; Tanaka et al. 1987; Westhoff and Herrmann 1988; Rock et al. 1987).

In this paper we extend our analysis of the barley *psb*D-*psb*C transcription unit and the 8 kbp DNA region between *psb*A gene and *psb*D-*psb*C. Earlier studies showed that at least eight different transcripts, ranging from 5.7 to 0.8 kb, hybridized to *psb*D-*psb*C (Berends et al. 1987). Some of these transcripts had their 5′ termini upstream of the DNA region previously analyzed. Therefore, in this study, we have cloned and sequenced the DNA region which contains the 5′ ends of these *psb*D-*psb*C RNAs. We located *trn*K(UUU), ORF504, *rps*16, *trn*Q(UUG), *psb*K, *psb*I, and *trn*S(GCU) upstream of the *psb*D-*psb*C transcription unit. Northern blots and analysis of RNA termini were done to further characterize the expression of these plastid genes.

Materials and methods

Plant growth. Barley (*Hordeum vulgare* L. Var. Morex) seeds were planted in vermiculite watered with half-strength Hoagland's nutrient solution. Seedlings were maintained in light-tight controlled environment chambers (at 23°C) in a light-tight room. After 4.5 days, dark-grown barley seedlings were either harvested or transferred to an illuminated chamber with a light intensity of 300 $\mu E/m^2$ for an additional 72 h. Manipulations of dark-grown barley plants were performed in complete darkenss except when necessary, a dim green safelight was used as described by Klein and Mullet (1986).

Plastid isolation. Intact plastids were isolated on percoll gradients as previously (Klein and Mullet 1986). Plastids were quantitated (in terms of plastid numbers per µl of suspension volume) by phase contrast microscopy.

Plastid nucleic and acid isolation. Total nucleic acid was isolated by phenol extraction from intact barley plastids as previously described (Mullet et al. 1985), and resuspended in H_2O on a per plastid basis.

DNA sequencing. Dideoxy sequencing reactions were performed on double-stranded plasmid DNA, using α^{32}S-ATP (from NEN), custom-made oligonucleotide primers, and the Sequenase kit (US Biochemical Corp. Cleveland, Ohio). DNA sequencing ladders were analyzed on 5% polyacrylamide-8.3 M urea gels (Maxam and Gilbert 1980). The DNA sequence presented in Fig. 2 was obtained by sequencing both DNA strands of the 6.25 kbp region to completion. DNA sequences were analyzed on computer using the IntelliGenetics program.

Northern blots. Total plastid nucleic acid was separated on 0.8% agarose/formaldehyde gels (according to Maniatis et al. 1982) and blotted onto Gene Screen filters by NEN's capillary transfer method. Double-stranded DNA probes were radiolabelled using a random primer labelling kit by BRL. Radiolabelled single-stranded RNA probes were generated from cloned plastid DNA fragments, using Stratagene's in vitro transcription kit. Northern filters were probed with the double-stranded DNA probes according to NEN's Gene Screen method II. For the single-stranded RNA probes, Stratagene's method for RNA-RNA hybridizations was employed; however, the hybridization temperature was changed to 55°C.

S_1 nuclease and primer extension assays. DNA fragments were 5′ or 3′ end labeled according to Maniatis et al. (1982). S_1 nuclease and primer extension assays were done using barley plastid RNA as previously described (Mullet et al. 1985). In some cases (see Fig. 6), unlabeled RNA synthesized from cloned plastid DNA (again using the Stratagene in vitro transcription kit) was used. Results were analyzed on 5% or 8% polyacrylamide-8.3 M urea gels (Maxam and Gilbert 1980).

Results

DNA sequence analysis of a 6.25 kbp region upstream of psbD-psbC

Barley plastid DNA encoding *psb*D, *psb*C, and DNA 3′ to these genes, was previously sequenced (Oliver and Poulson 1984; Neumann 1988; Efimov et al. 1988a). Northern blot analysis (Berends et al. 1987), showed that the *psb*D-*psb*C transcription unit includes an unidentified open reading frame of 62 amino acids (ORF62), a divergent *trn*S(UGA) gene, and a convergent *trn*G(GCC) gene downstream of *psb*D-*psb*C (Fig. 1). DNA sequence analysis of a 2.2 kbp *Eco*RI fragment located approximately 7.5 kbp upstream of *psb*D-*psb*C revealed several divergent genes, including *psb*A, the 3′ exon of *trn*K(UUU), and a partial open reading frame encoding 150 C-terminal amino acids (Boyer and Mullet 1988; Efimov et al. 1988b). In the present study we have sequenced a 6.25 kbp barley ctDNA region (Fragment K in Fig. 1, located between two *Eco*RI sites), bridging the gap between the partial open reading frame and the *psb*D-*psb*C genes.

The 6.25 kbp DNA sequence is presented in Fig. 2. Computer analysis revealed numerous open reading frames (ORFs) on both DNA strands, ranging in size from 2 to 354 amino acid codons. ORFs homologous to previously identified genes on other plastid genomes have been marked in Figs. 1 and 2. Of the remaining ORFs, many were small and were not transcribed. ORF41 has been marked in Figs. 1 and 2 because it may be cotranscribed with *psb*I; both *psb*I and ORF41 overlap ORF44 (also marked in Figs. 1 and 2) which is encoded on the opposite DNA strand. The largest ORF was located near *psb*A (Fig. 2). This ORF is encoded on the DNA strand opposite to *psb*D-*psb*C and its predicted 345 N-terminal acid sequence joins an adjacent C-terminal amino acid sequence to form ORF504. Identification of the *trn*K(UUU) 5′ exon in Fig. 2 places ORF 504 within the *trn*K(UUU) intron. 566 bp upstream of *trn*K(UUU) on the same DNA strand lies the 3′ exon of *rps*16, which is separated from its 5′ exon by a 790 bp intron. Located between *rps*16 and *psb*D, an ORF encoding 61 or 64 amino acids was found to resemble *psb*K in tobacco (ORF98; Shinozaki et al. 1986b). *psb*I (encoding 36 amino acids) was located 413 bp downstream of *psb*K, on the same DNA strand. In barley, *psb*I overlaps an ORF41 of unknown function. Interestingly, a third open reading frame (ORF44) overlaps both *psb*I and ORF41 on the opposite DNA strand.

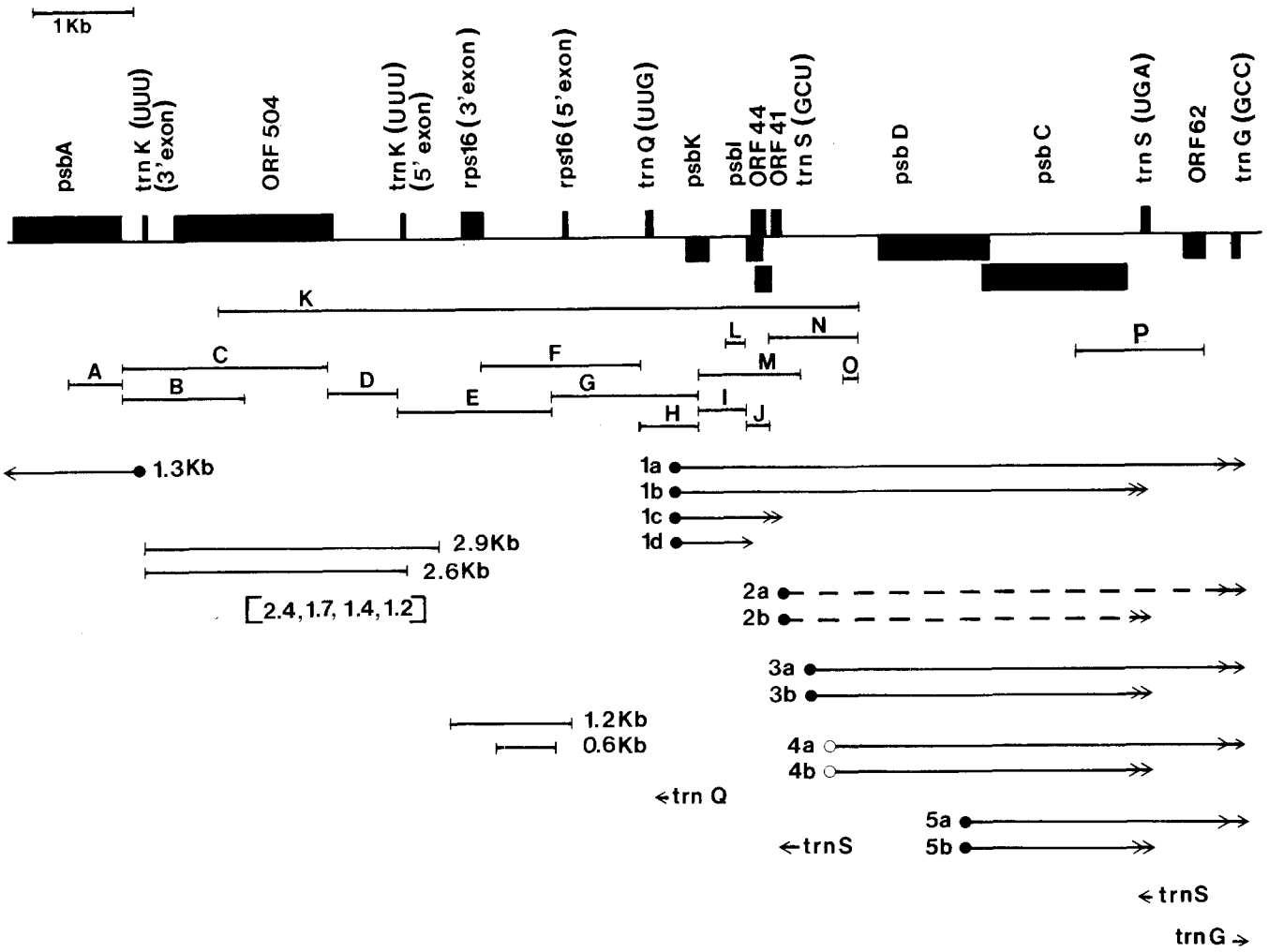


Fig. 1. Physical and transcript map of a 12.5 kbp barley ctDNA region. *Solid boxes* denote open reading frames, tRNAs, or exons. Genes above the dividing line are encoded on the same DNA strand from right to left; genes below the line are encoded on the opposite strand from left to right. *Fragments A through J* represent the double-stranded DNA probes used in Fig. 3. The DNA sequence of fragment K is presented in Fig. 2. Fragments L through P were used in Fig. 5 through 7. *Solid lines* below the gene map represent those transcripts which have been visualized on Northern blots. *Dotted lines* represent those transcripts which were identified by transcript 5'-end analysis but which could not be clearly resolved on Northern blots. *Small arrows* indicate tRNAs which are expressed in etioplasts of 4.5-day old, dark-grown barley. *Arrow heads* mark confirmed 3' ends. Confirmed 5' ends are marked by *solid circles* (for transcripts present in barley etioplasts) or *open circles* (for light-induced transcripts). Several low abundance RNAs were identified in the *trn*K-ORF504 region but were difficult to map. The sizes of these RNAs are shown in brackets. Approximate transcript sizes are indicated. Transcripts from the *psb*D-*psb*C transcription unit have been labelled according to their 5' and 3' ends. Their approximate sizes (determined by Northern blotting) are as follows: 1a (5.7 kb); 1b (4.8 kb); 1c (1.0 kb); 1d (0.8 kb); 3a (4.1 kb); 3b (3.2 kb); 4a (4.0 kb); 4b (3.1 kb); 5a (2.6 kb); and 5b (1.7 kb)

Northern blot analysis

In previous publications (Berends et al. 1987; Gamble et al. 1988), we described eight *psb*D-*psb*C transcripts, six of which were present in barley etioplasts (transcripts 1a, 1b, 3a, 3b, 5a, and 5b in Fig. 1), and two of which accumulate upon illumination (transcripts 4a and 4b in Fig. 1). Six *psb*D-*psb*C transcripts, including the light-induced transcripts, were mapped downstream of the 6.25 kbp region described above. These transcripts (3a, 3b, 4a, 4b, 5a, and 5b) can be divided into three groups, each of which shares a common 5' end (indicated by numbers). A fourth 5' end downstream of the 6.25 kbp region belongs to transcripts 2a and 2b, which will be discussed in a later section (see Fig. 1). For each 5' end there are two clusters of 3' ends (indicated by letters) that were mapped around the divergent *trn*S(UGA) and convergent *trn*G(GCC) genes downstream of *psb*C. These 3' ends are shared also by the two longest *psb*D-*psb*C transcripts (transcripts 1a and 1b) present in barley etioplasts. 5' end-mapping data for the latter transcripts however was incomplete, since these transcripts extended beyond the cloned DNA fragment under investigation, into the 6.25 kbp region. A 1.0 kb transcript (transcript 1c in Fig. 1), whose 3' end was mapped upstream of *psb*D-*psb*C (Berends et al. 1987), also seemed to originate within the 6.25 kbp region. To complete the *psb*D-*psb*C transcript map, and as a preliminary step towards studying the expression of the newly

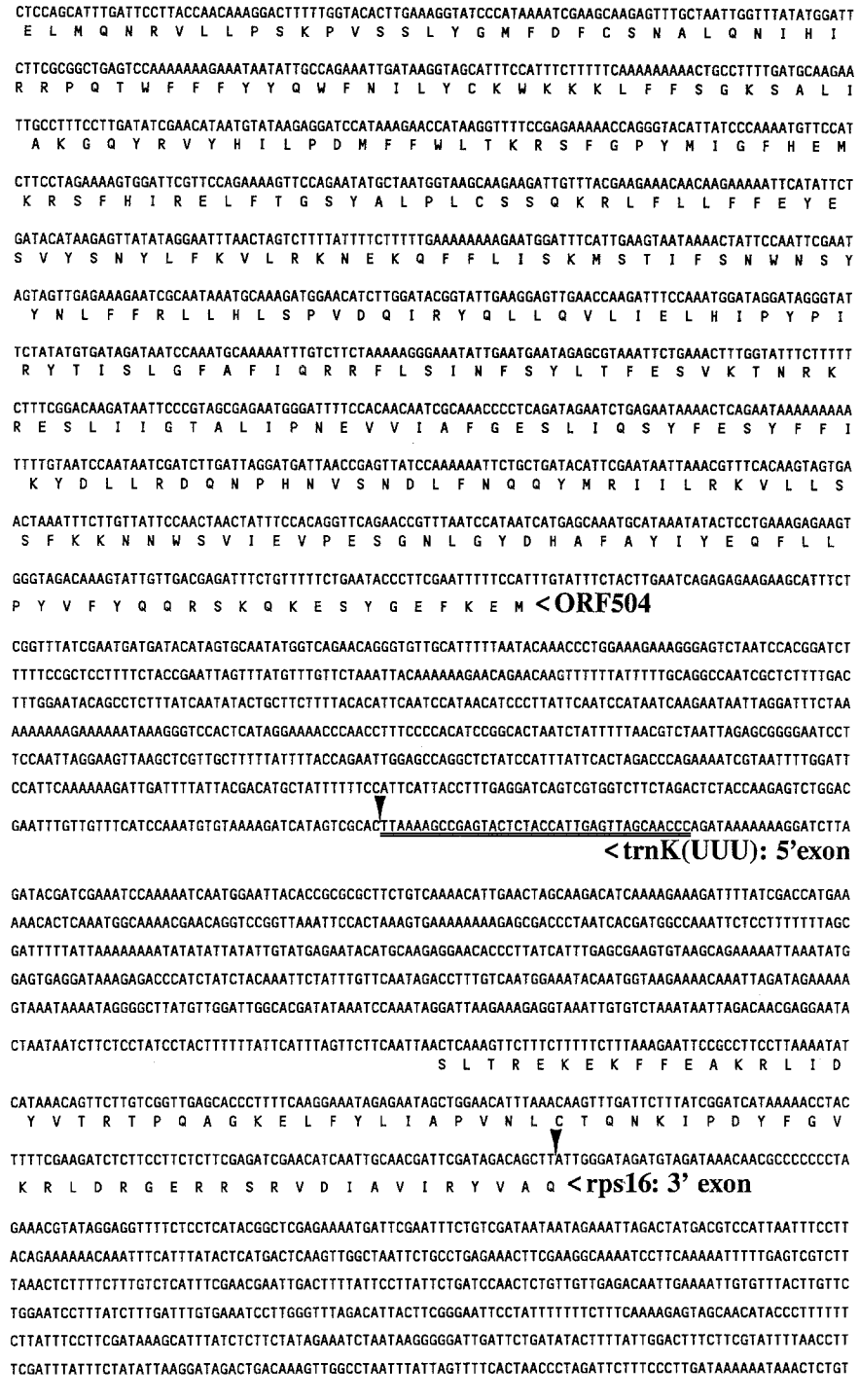
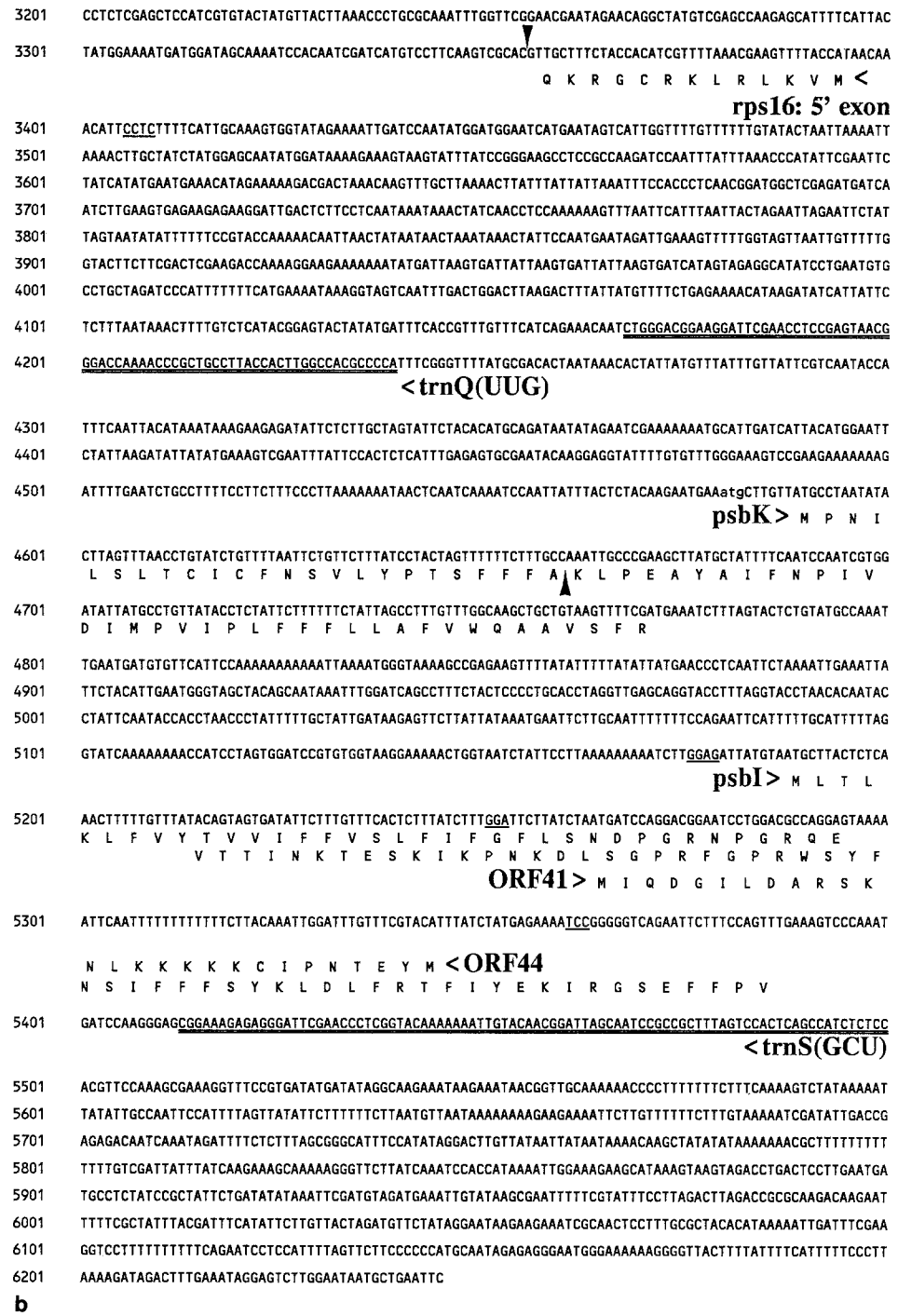


Fig. 2. DNA sequence analysis of the 6.25 kbp region between *psb*A and *psb*D-*psb*C on the barley plastid genome (fragment K in Fig. 1). The sequence is presented in a 5′ to 3′ orientation relative to *psb*D-*psb*C. Genes or exons which encode tRNAs have been *underlined twice*. Predicted amino acid sequences of open reading frames described in the text are shown below the DNA sequence. Only the 354 N-terminal amino acid sequence of ORF504 is shown here; the remaining 150 C-terminal amino acid sequence was presented in a previous publication (Boyer and Mullet 1988). A non-conserved, alternative ATG methionine initiation codon for the protein encoded by *psb*K has been printed in *lower case letters*. The *arrow below the psb*K-*derived amino acid sequence* shows a proposed protein processing site (Murata et al. 1988). RNA splice junctions are indicated by *arrows above the DNA sequence*. Putative ribosome binding sequences are *underlined*



b

sequenced *trn*K(UUU), ORF504, and *rps*16 genes upstream of *psb*D-*psb*C, we extended our Northern blot analysis to include the 6.25 kbp region.

Identical Northern blots of plastid RNA from 4.5-day old, dark-grown barley seedlings were probed with radiolabeled, double-stranded DNA probes A through J (see Fig. 1). The resulting autoradiograms are shown in Fig. 3, where lanes A–J correspond to probes A–J. Starting with probe A (a 0.51 kbp *Hin*dIII-*Bgl*II fragment), a single 1.3 kb *psb*A transcript could be seen. This transcript had been previously mapped and was found to initiate 80 bp upstream of the *psb*A open reading frame, in a direction divergent from the barley *psb*D-*psb*C genes (Boyer and Mullet, 1988; see Fig. 1). The same 1.3 kb RNA hybridized to probes B (a 1.21 kbp *Bgl*II-*Eco*RI fragment) and C (a 2.03 kbp *Bgl*II-*Acc*I fragment), but not with D (a 0.67 kbp *Acc*I-*Xba*I fragment). In the *trn*K/ORF504 region probe D did hybridize to at least nine other transcripts, including three abundant closely migrating transcripts approximately 2.9, 2.6 and 2.4 kb in size. The 2.9 kb and 2.6 kb transcripts extended into fragments B and C, and could be seen as faint bands with probe E (a 1.53 kbp *Xba*I-*Sac*I

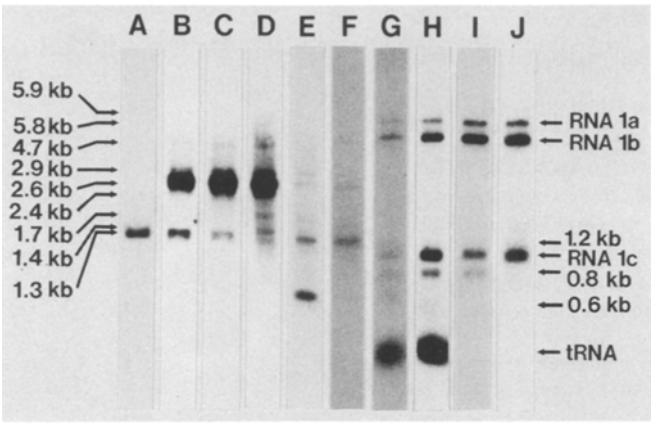


Fig. 3. Northern blot analysis of the barley ctDNA region containing *psb*A, *trn*K(UUU), ORF504, *rps*16, *trn*Q(UUG), *psb*K, ORF41, *psb*I, ORF44, and *trn*S(GCU). *Lanes A through J* contained barley plastid RNA from 4.5-day old, dark-grown seedlings probed with double-stranded DNA probes A through J (diagrammed in Fig. 1). Approximate transcript sizes are indicated. Previously described transcripts in the *psb*D-*psb*C transcription unit (Berends et al. 1987) are labelled as described in the text and in Fig. 1

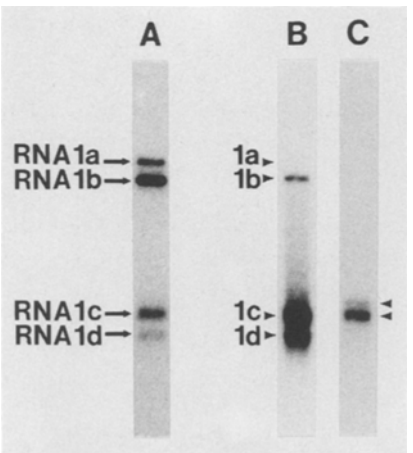


Fig. 4. Transcriptional orientation of *psb*D-*psb*C transcripts 1a, 1b, 1c and 1d. *Lanes A, B and C* of the Northern blot contain plastid RNA from 4.5-day old, dark-grown barley seedlings. Lane A was probed with double-stranded DNA fragment I in Fig. 1. Lanes B and C were probed with complementary single-stranded RNA probes derived from fragment

fragment). A low abundance 2.6 kb RNA of undetermined origin was also detected with probe F (a 1.58 kbp *Bgl*II-*Eco*RV fragment). Based on approximate size and position, the 2.9 and 2.6 kb transcripts could represent unspliced *trn*K/ORF504 messages. It is difficult to say whether the 2.4 kb transcript extends into fragments B and C. Also in lane D were three very low abundance transcripts of approximately 4.7, 5.8 and 5.9 kb. The 4.7 kb band extended into fragments B and C, while the 5.8 and 5.9 bands could not be detected in any other lanes. Considering their low abundance, compared to other *trn*K transcripts and the *psb*A message, and their large size, which may hamper capillary transfer to Northern filters, it is possible that these large transcripts extend through the *psb*A region. Such transcripts could act as precursors to the mature $tRNA^{lys}$. Probe D in the *trn*K region also hybridized to three smaller transcripts (of approximately 1.7, 1.4 and 1.2 kb), which were also low in abundance. In some blots, upon long exposure, an RNA of tRNA size was detected with probes B, C and E. The lack of a strong signal from $tRNA^{lys}$could be due to low abundance, RNA secondary structure which results in poor hybridization or poor retention of this RNA on our blots. A 1.2 kb RNA was detected with fragments E, F and G (a 1.46 kb *Sac*I-*Hin*dIII fragment), but not with fragment H (a 0.58 kbp *Eco*RV-*Hin*dIII fragment). Fragments E, F and G are part of the *rps*16 region. It is possible that this transcript represents an unspliced precursor to *rsp*16 or else a splicing product. In addition, a small (0.6 kb) transcript hybridized strongly to probe E and weakly to probes F and G. This transcript could encode the *rps*16 open reading frame. Fragments G and H also hybridized to four transcripts, three of which (RNAs 1a, 1b and 1c) have been previously mapped (Berends et al. 1987). The fourth RNA (a 0.8 kb transcript) and a transcript of tRNA-size were detected by probe G. The tRNA-size transcript was also detected by probe H, and could represent $tRNA^{gln}$ encoded by *trn*Q(UUG). The absence of RNAs 1a, 1b, 1c and the 0.8 kb transcripts in lane F indicates that these transcripts have a terminus within fragment H. Transcripts 1a, 1b and 1c extend through fragments I (a 0.45 kbp *Hin*dIII-*Bam*HI fragment) and J (a 0.25 kbp *Bam*HI-*Eco*RI fragment). The 0.8 kb transcript on the other hand could not be detected in lane J.

Mapping of barley psbD-psbC transcripts 1a-1d

The 0.8 kb transcript described above may be cotranscribed as part of the barley *psb*D-*psb*C transcription unit. To investigate this possibility, we first determined its orientation by Northern blot analysis, using single-stranded RNA probes. For comparison, lane A of Fig. 4 shows an autoradiogram of a Northern blot using double-stranded probe I (see Figs. 1 and 3). Of the four bands in lane A, the top three can be attributed to previously mapped transcripts, 1a, 1b, and 1c of the *psb*D-*psb*C transcription unit (Berends et al. 1987). To determine the orientation of the transcript giving rise to the smallest (0.8 kb) band, we probed identical Northern blots with single-stranded RNA probes derived from both strands of DNA region I (lanes B and C Fig. 4). In lane B, all four RNAs could be seen, indicating that a 0.8 kb RNA (named transcript 1d) is transcribed from the same DNA strand as transcripts 1a, 1b and 1c (note that transcript 1a could only be seen upon long exposure). The complementary RNA probe (in lane C) did not hybridize to transcripts 1a, 1b or 1d. It did, however, hybridize to two transcripts approximately 1.1 and 1.0 kb in size. The 1.1 kb transcript was of minor abundance compared to the 1.0 kb transcript, which may explain its apparent absence in lane A. Since the 1.0 kb transcript comigrated exactly with RNA 1c in lane B, they would be represented by a single band in

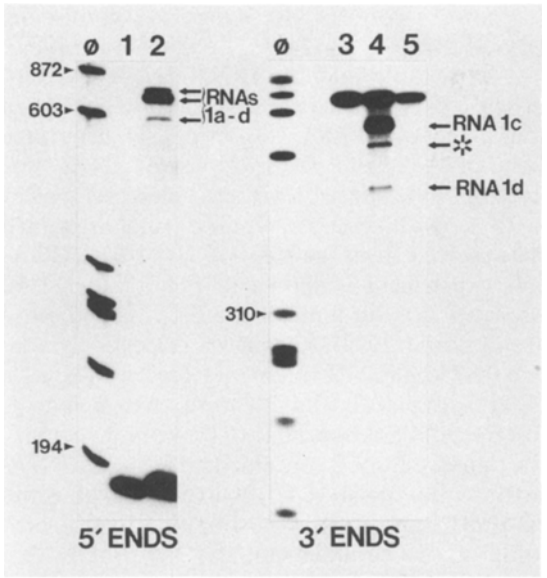


Fig. 5. Mapping of 5′ and 3′ termini of *psb*D-*psb*C transcripts 1a–1d. ϕ refers to end-labeled ϕX174/*Hae*III fragments used as molecular weight markers. *Lane 1* contains the 187 bp *Sau*3A-*Bam*HI primer used in the extension reaction (on barley etioplast RNA) in *lane 2*. *Lane 3* contains the 1.02 kbp *Hin*dIII-*Cla*I fragment used as a 3′ S_1 nuclease probe in *lanes 4* (on barley etioplast RNA) *and 5* (no RNA added)

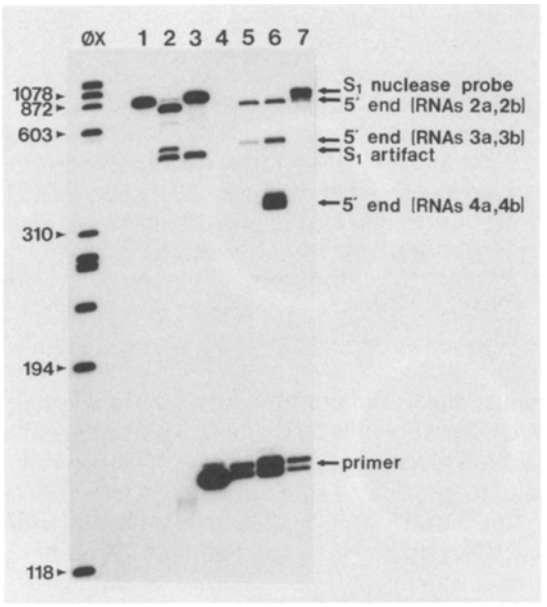


Fig. 6. Verification of *psb*D-*psb*C S_1 nuclease and primer extension signals. *Lane 1* shows the 0.87 kb end-labelled *Eco*RI fragment used as S_1 nuclease probe in *lanes 2 and 3*. *Lane 4* shows the 135 bp *Sau*961-*Eco*RI fragment (end-labelled only at the *Eco*RI site) used as primer in the primer extension assays in *lanes 5–7*. Control assays were run on plastid RNA from 4.5-day old, dark-grown barley with (lane 6) or without (lanes 2 and 5) an additional 72 h of illumination. Synthetic *psb*D-*psb*C RNA derived from the 0.87 kb *Eco*RI fragment was used in lanes 3 and 7. The presence of 5′ end signals in lanes 3 and 7 indicates an artifact

lane A. Although the 5′ and 3′ termini of these divergent transcripts have not yet been determined, it is likely that such transcripts encode ORF44 on the DNA strand opposite to *psb*D-*psb*C.

Northern blot analysis indicated that transcripts 1a, 1b and 1c may initiate at a common 5′ end. The orientation and approximate position of transcript 1d suggest it may also originate from this 5′ end. To check this possibility, and to further map transcript 1d, we performed a series of S_1 nuclease and primer extension analyses. The results are shown in Fig. 5. Primer extension of 187 bp *Sau*3A-*Bam*HI DNA (Fig. 1, fragment L), labeled at the *Bam*HI site, revealed three 5′ end signals (Fig. 5, lane 2). The major (middle) signal mapped approximately 720 bp from the *Bam*HI site, about 2.0 kb upstream of *psb*D. The two minor signals mapped respectively to about 750 and 600 bp upstream of the *Bam*HI site. Although it was not possible to assign individual 5′ ends to transcripts 1a–1d, this analysis showed that all four transcripts initiate within 150 bp of each other.

The 3′ end of transcript 1d was mapped with a 1.02 kb *Hin*dIII-*Cla*I DNA (Fig. 1, fragment M) 3′ end labeled at the *Hin*dIII site (Fig. 5, lane 3). S_1 nuclease treatment of the probe alone did not result in any additional bands (lane 5). S_1 nuclease protection assays of barley etioplast RNA, however, resulted in three 3′ end signals, approximately 800, 700 and 480 bp downstream of the *Hin*dIII site (lane 4). The major band (800 bp) corresponded to the 3′ end of transcript 1c. The starred band (700 bp) did not belong to any known RNA. This 3′ was not detected in our earlier mapping experiments in this region (Berends et al. 1987), possibly due to its low abundance. As one might expect from the band intensities in lane B of the Northern blot in Fig. 4, the 480 bp band (3′ end of transcript 1d) was also minor compared to the 3′ end signal of transcript 1c. The mapping results for transcripts 1a–1d are summarized in Fig. 1.

Identification of the 5′ terminus of RNAs 2a and 2b

A strong primer extension/S_1 nuclease signal was previously noted approximately 260 bp upstream of the 5′ end of RNAs 3a and 3b (Berends et al. 1987). To investigate the nature of signal in detail, and to verify other primer extension/S_1 nuclease signals in this region, we performed the experiment illustrated in Fig. 6. In this experiment, a synthetic *psb*D-*psb*C RNA was prepared from the 0.87 kb *Eco*RI fragment (DNA region N, Fig. 1) which also contains *trn*S(GCU). The synthetic RNA (lanes 3 and 7) and plastid RNA from 4.5-day old, dark-grown barley (lanes 2 and 5), or plastid RNA from similar plants illuminated for an additional 72 h (lane 6), was analyzed by S_1 nuclease (lanes 2 and 3) or primer extension assays (lanes 5, 6, and 7). Lane 1 shows the 0.87 kb *Eco*RI fragment used as the S_1 nuclease probe (Fig. 6). A 135 bp *Sau*961-*Eco*RI fragment (DNA region O, Fig. 1), labelled at its *Eco*RI site, served as the primer (lane 4, Fig. 6). A strong signal indicated in Fig. 6 as the 5′-end of RNAs 2a, 2b, was present in the in vivo S_1 nuclease (lane 2) and primer extension assays (lanes 5 and 6). A minor signal in

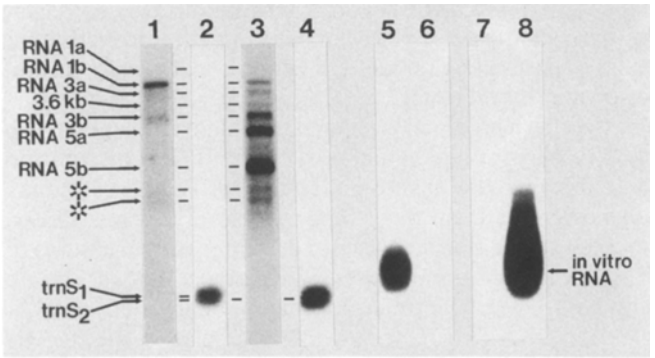


Fig. 7. Expression of *trn*S(GCU) and *trn*S(UGA) in plastids of 4.5-day old, dark-grown barley. *Lanes 1–4* of the Northern blot contain in vivo barley RNA. *Lanes 5 and 7* contain an in vitro-synthesized RNA which includes *trn*S(GCU), and *lanes 6 and 8* contain in vitro-synthesized RNA which includes *trn*S(UGA). Lane 1 was probed with a radiolabelled synthetic RNA homologous to *psb*D-*psb*C transcripts and "anti-sense" *trn*S(GCU). Similarly, lane 3 was probed with a synthetic RNA probe homologous to *psb*D-*psb*C transcripts and "anti-sense" *trn*S(UGA). Lanes 2, 5 and 6 were probed with a synthetic RNA probe homologous to *trn*S(GCU). Lanes 4, 7 and 8 were probed with a synthetic RNA probe homologous to *trn*S(UGA)

this region also appeared in the primer extension assay on the synthetic RNA (lane 7). Since a synthetic 0.87 kb RNA should give rise to S_1 nuclease or primer extension signals only from the RNA termini, this suggests that this RNA can cause termination of reverse transcriptase at a low frequency at this site. In the S_1 nuclease assay (lane 3), the same synthetic RNA preparation did not generate a signal, showing that the corresponding S_1 nuclease signal in lane 2 in vivo assay) must be due to an RNA 5′ end. At the same time, the absence of the unidentified signal in lane 3 proves that the synthetic RNA preparation is free of 5′ ends which could otherwise cause the signal to appear in lane 7. As a positive control for identifying S_1 artifacts by this method, the S_1 artifact in lane 2, previously identified (Berends et al. 1987) because of its absence in the corresponding in vivo primer extension assays (lanes 4 and 5), did appear in lane 3. Conspicuously absent from S_1 nuclease (lane 3) or primer extension (lane 7) assays with the synthetic RNA were the 5′ end signals of transcripts 3a (seen in lanes 2, 5, and 6), and of transcripts 4a and 4b (seen in lane 6), confirming that these signals arise from RNA termini. Therefore, we conclude that the signal generated upstream from 3a and 3b is generated by an RNA 5′ end. Since the only 3′ ends which have been found downstream of this 5′ end signal are those shared by the other *psb*D-*psb*C transcripts, we have tentatively assigned the 5′ end to transcripts 2a and 2b in Fig. 1.

trnS(GCU) and trnS(UGA) expression in barley etioplasts

Two tRNA genes, *trn*S(GCU) and *trn*S(UGA, are encoded within the barley *psb*D-*psb*C transcription unit but on the opposite DNA strand. Northern blot analysis with double-stranded DNA probes has shown that tRNA-size transcripts arise from DNA regions containing *trn*S(GCU) and *trn*S(UGA) (Fig. 3, Berends et al. 1987). However, it is possible that the tRNA-size RNAs represent anti-sense tRNA structures, which may have been excised from longer *psb*D-*psb*C transcripts. To determine from which DNA strand the tRNA-size RNAs was synthesized, we hybridized Northern blots of barley etioplast RNA with complementary single-stranded RNA probes derived from the *trn*S(GCU) or *trn*S(UGA) region. This experiment is shown in Fig. 7 (lanes 1-4). The RNA probes used in lanes 1 and 2 contained sense and anti-sense *trn*S(GCU) sequences respectively, derived from a 0.87 kbp *Eco*RI fragment (Region N, Fig. 1). *Psb*D-*psb*C transcripts 1b and 3b were seen in lane 1. These transcripts did not hybridize to the complementary RNA probe (lane 2). Lane 2 contained an RNA of tRNA-size, indicating that *trn*S(GCU) is expressed at times when *psb*D-*psb*C transcripts in this region are also produced. Similarly, two complementary RNA probes were derived from a 1.26 kbp *Hin*dIII-*Eco*RI fragment (region P, Fig. 1) that contained *trn*S(UGA). Six *psb*D-*psb*C transcripts (RNAs 1a, 1b, 3a, 3b, 5a, and 5b) appear in lane 3. Bands of 3.6 kb and of approximate rRNA-size (starred, lane 3) were also observed (Berends et al. 1987; Oliver and Poulson 1984). However, no RNAs of tRNA-size were detected in lane 3. An RNA of tRNA-size was detected with the complementary RNA probe in lane 4, suggesting that *trn*S(UGA) is also a functional gene. Since both genes encode $tRNA^{ser}$, we tested for cross-hybridization by synthesizing non-labelled versions of the *trn*S(GCU) and *trn*S(UGA) probes used in lanes 1 and 3 respectively. Northern blots of the in vitro-synthesized RNAs containing *trn*S(GCU) (lanes 5 and 7) and *trn*S(UGA) (lanes 6 and 8) were then probed with radiolabelled probes capable of hybridizing to $tRNA^{ser(GCU)}$ (used in lanes 2, 5 and 6) or $tRNA^{ser(UGA)}$ (used in lanes 4, 7 and 8). Even after overexposure of the Northern blots, no cross-hybridization could be detected, indicating that the bands seen in lanes 2 and 4 represent $tRNA^{ser(GCU)}$ and $tRNA^{ser(UGA)}$ respectively.

Discussion

DNA sequence analysis of the 6.25 kbp region between the *psb*A and *psb*D-*psb*C genes on the barley plastid genome revealed several open reading frames (ORFs) or exons homologous to previously identified genes (see Fig. 1). The barley *trn*K(UUU) gene lies 230 bp upstream of *psb*A on the same DNA strand (Boyer and Mullet 1988, and this study). Its 2.48 kbp intron contains a large ORF (ORF504), which resembles ORF509A of the tobacco plastid genome (Sugita et al. 1985). ORF509A is thought to encode an RNA maturase involved in RNA splicing. Although most higher plants contain a large ORF within the *trn*K intron, only 52% of the amino acids predicted by each DNA sequence are conserved between ORF504 in barley and ORF509A in tobacco. Similarly, only 66% of overall homology has been found between the predicted amino acid sequences of ORF509A in tobacco and ORF524 in mustard (Neuhaus and Link 1987). A 103-

amino acid sequence near the C-terminus of the protein encoded by ORF524 of mustard has been shown to be structurally related to maturase-like polypeptides of mitochondrial introns (Neuhaus and Link 1987). Comparison of the predicted amino acid sequences in the equivalent regions of ORF504 of barley (amino acids 354–457; Boyer and Mullet 1988; this study) to ORF509A of tobacco (amino acids 354–457; Sugita et al. 1985) raises the homology from 52% to 79%. The split *rps*16 gene upstream of *trn*K(UUU) is overall more highly conserved among plastid genomes; *rps*16 of tobacco (Shinozaki et al. 1986a) and mustard (Neuhaus et al. 1989) exhibited 85% and 77% of predicted amino acid sequence homology, respectively, to that of barley. Northern analysis indicates that the *trn*K(UUU) and *rps*16 regions are transcribed in etioplasts of 4.5-day old, dark-grown barley.

Evolutionary DNA rearrangements have been proposed to explain the close proximity and divergent orientation of the *psb*A and *psb*D-*psb*C genes in monocots (Quigley and Weil 1985; Courtice et al. 1985; Berends et al. 1987). Such DNA rearrangements have placed the *psb*D-*psb*C genes immediately downstream of a conserved DNA region containing *trn*Q(UUG), *psb*K, *psb*I and *trn*S(GCU). Genes *trn*Q(UUG) and *trn*S(GCU) are located on the DNA strand opposite of *psb*D-*psb*C, while *psb*K and *psb*I are on the same DNA strand. Interestingly, groups of tRNA genes surround the *psb*D-*psb*C and the *psb*K-*psb*I genes in both dicots and monocots. tRNA genes have been proposed to act as insertion sites during DNA rearrangements in prokaryotes (Reiter et al. 1989). It is possible that tRNA genes have served a similar function during plastid evolution. In support of this theory, Howe et al. (1988) have noted several short repeats in the *trn*S(GCU)-*psb*D intergenic region of wheat, and in the *trn*G(UCC)-*trn*T intergenic regions of barley and maize. Close inspection of the barley *trn*S(GCU)-*psb*D intergenic region (Efimov et al. 1988a; Fig. 2) reveals their presence here as well. Thought to be indicative of DNA rearrangements, these repeats are absent in the ancestral genomes of tobacco, spinach and pea (Howe et al. 1988).

In tobacco (a dicot), *psb*K is present as ORF98, proposed to start at a valine (GTG) initiation codon (Deno and Sugiura 1983; Shinozaki et al. 1986b). N-terminal amino acid sequencing of a 4.3 kDa PSII protein from spinach, however, has led to the proposal of a post-translational site to yield a mature *psb*K-encoded protein of 37 amino acids (Murata et al. 1988). The proposed GTG initiation codon of ORF98/*psb*K in tobacco is a GTT (valine) codon in barley, and this codon is followed by stop codons in all three reading frames. Instead, translation initiation at two downstream methionine (ATG) initiation codons (in frame with each other) could result in a 64 or 61-amino acid protein respectively. Either precursor protein could be processed at the putative conserved processing site (Fig. 2, arrow) to yield a mature 37-amino acid protein. Interestingly, only the second ATG codon is conserved in tobacco and mustard (Neuhaus 1989), suggesting that the primary translation product could span 61, rather than 64, amino acids. Neither ATG codon appears to be preceded by a strong ribosome binding site. However, there are several examples of plastid genes known to be expressed at the protein level despite lack of a putative ribosome binding sequence (Herrmann et al. 1985). Comparison of the DNA-deduced amino acid sequences in tobacco (Deno and Sugiura 1983) or mustard (Neuhaus 1989) to that of barley shows a high degree of conservation (95% each) for the proposed mature *psb*K protein.

*Psb*I is located 413 bp downstream from *psb*K. This gene encodes a 36-amino acid protein which has 97% of homology to the protein encoded by tobacco (Shinozaki et al. 1986b; Ikeuchi and Inoue 1988) or wheat (Howe et al. 1988). the *psb*I ORF is preceded by a putative ribosome binding site (GGAG). ORF41 overlaps *psb*I by 35 bp on the same DNA strand. On the opposite DNA strand, ORF44 overlaps both *psb*I and ORF41. ORFs which overlap *psb*I are found in other plastid genomes as well, but differ in predicted amino acid length (90 and 33 tobacco, Shinozaki et al. 1986b; 45 and 28 wheat, Howe et al. 1988) and are poorly conserved at the amino acid level. Computer-assisted homology searches did not indicate a significant resemblance of the ORF41 and ORF44-deduced amino acid sequences to other proteins. However, weak putative ribosome binding sites (GGA) precede both ORF41 and ORF44. Given that ORF41 and ORF44 are transcribed in etioplasts of 4.5 day-old, dark-grown barley, it will be interesting to see if ORF41 and ORF44 encode functional proteins.

Previous studies (Berends et al. 1987; Gamble et al. 1988) showed that at least eight transcripts (mRNAs 1a, 1b, 3a, 3b, 4a, 4b, 5a and 5b; see Fig. 1) hybridize to *psb*D-*psb*C on the barley plastid genome. Of these transcripts, the 5′ end of mRNAs 3a through 5b were mapped to completion. We have now extended the *psb*D-*psb*C transcript map to include transcripts 2a and 2b (see Fig. 1). 5′ end-analysis revealed that transcripts 1a, 1b, 1c and 1d originate from a 150 bp region upstream of *psb*K (see Fig. 1). Transcripts 1a and 1b extend through *psb*K, *psb*I and ORF41, ending beyond *psb*C. Transcript 1c ends around the divergent *trn*S(GCU) gene upstream of *psb*D, while transcript 1d appears to encode *psb*K only, stopping just short of *psb*I. In barley PSII, genes *psb*K, *psb*I, *psb*D, and *psb*C are all part of the same transcription unit, possibly allowing for tighter coregulation than in tobacco or spinach.

Conservation of transcription units among different plastid genomes suggests that coregulation of the genes within the transcription unit is important. Furthermore, the position of transcription units on the plastid genome may be significant as well. Northern blot analysis indicates that the *trn*Q(UUG) gene immediately upstream of the barley *psb*D-*psb*C transcription unit is transcribed in etioplasts of 4.5-day old, dark-grown barley. This implies that two divergent promoters lie in close proximity to each other (within about 180 bp). Adjacent, divergent promoters of the *atp*B and *rbc*L genes on the maize chloroplast genome (within a 159 bp sequence) have been shown to interact in in vitro transcription experiments (Hanley-Bowdoin and Chua 1989). Binding of RNA polymerase to the *rbc*L promoter interferes with *atp*B promoter function. Although the *psb*D-*psb*C transcription unit of spinach and tobacco remain to be elucidated, it is interesting

to note that even in dicots, three conserved tRNA genes upstream of *psb*D-*psb*C are also divergent (Shinozaki et al. 1986b; Holschuh et al. 1984).

Several divergent transcripts arising from within the *psb*D-*psb*C transcription unit were noted in etioplasts of 4.5-day old, dark-grown barley. Northern blot analysis using single-stranded RNA probes revealed two transcripts of approximately 1.1 and and 1.0 kb which were complementary to *psb*D-*psb*C transcripts 1a through 1d. These transcripts could encode ORF44, which overlaps *psb*I and ORF41 on the opposite DNA strand. Two tRNAs with serine anticodons [encoded by the *trn*S(GCU) and *trn*S(UGA) genes] were also produced from the opposite DNA strand within the *psb*D-*psb*C transcription unit. Divergent transcription within an operon is not unique to the *psb*D-*psb*C transcription unit. It has also been noted in the *psb*B-*psb*H-*pet*B-*pet*D transcription unit in livewort, where transcription of a divergent ORF43 in the *psb*B-*psb*H spacer region results in an mRNA complementary to part of the *psb*B primary transcript (Kohchi et al. 1988). In prokaryotes, binding of anti-sense RNAs to target mRNAs has been proposed to cause inhibition of ribosome binding, premature termination of target RNA transcription, interference with the formation of target RNA secondary structures needed for processing, or prevention of read-through transcription (for a review see Simons and Kleckner 1988). Interestingly, the 3′ ends of transcripts 1c were mapped around *trn*S(GCU), and the 3′ ends of transcripts 1b, 2b, 3b, 4b, and 5b lie near *trn*S(UGA). It is conceivable that, aside from $tRNA^{ser}$ production, these *trn*S genes are involved in generating 3′ ends of complementary transcripts. Anti-sense *trn*S sequences, for instance, could form secondary structures in DNA or RNA, serving as transcription terminators or RNA processing sites respectively. Alternatively, simultaneous transcription of divergent genes on the same DNA template could cause transcription termination. In any case, divergent transcription within the barley *psb*D-*psb*C transcription unit may modulate gene expression from this DNA region.

Acknowledgements. The authors thank Mrs. Sharyll Pressley for her help in preparing the manuscript.

References

Alt J, Westhoff P, Sears BB, Nelson N, Hurt E, Hauska G, Herrmann RG (1983) EMBO J 2:979–986
Bedbrook JR, Link G, Coen DM, Bogorad L, Rich A (1978) Proc Natl Acad Sci USA 75:3060–3064
Bendich AJ (1987) BioEssays 6:279–282
Berends T, Gamble PE, Mullet JE (1987) Nucleic Acids Res 15:5217–5240
Boyer SK, Mullet JE (1988) Nucleic Acids Res 16:8184
Courtice GRM, Bowman CM, Byer TA, Gray JC (1985) Curr Genet 10:329–333
Deno H, Sugiura M (1983) Nucleic Acids Res 11:8407–8414
Efimov VA, Andreeva AV, Reverdatto SV, Chakhmakhcheva OG (1988a) Nucleic Acids Res 16:5680
Efimov VA, Andreeva AV, Reverdatto SV, Jung R, Chakhmakhcheva OC (1988b) Nucleic Acids Res 16:5685
Gamble PE, Sexton TB, Mullet JE (1988) EMBO J 7:1289–1297
Hanley-Bowdoin L, Chua N-H (1989) Mol Gen Genet 215:217–224
Herrmann RG, Westhoff P, Alt J, Tittgen J, Nelson N (1985) In: van Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular form and function of the plant genome. Plenum Publishing Corporation, New York, pp233–256
Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C, Meng B, Li Y, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) Mol Gen Genet 217:185–194
Holschuh K, Bottomley W, Whitfeld PR (1984) Plant Mol Biol 3:313–317
Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) Curr Genet 13:343–349
Ikeuchi M, Inoue Y (1988) FEBS Lett 241:99–104
Klein RR, Mullet JE (1986) J Biol Chem 261:11138–11145
Kohchi T, Yoshida T, Komano T, Ohyama K (1988) EMBO J 7:885–891
Link G (1982) Planta 154:81–86
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Maxam AM, Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical cleavages, In: Grossman L, Moldave K (eds) Methods in enzymology, vol 65. Academic Press, New York, pp499–560
Moller BL, Schneller HV, Okkels JS, Koch B, Anderson B, Nielsen HL, Olsen I, Halkier BA, Hoy PB (1990) Proceedings of the VIIIth International Congress on Photosynthesis. Kluwer Academic Publishers (in press)
Mullet JE, Orozco EM, Chua N-H (1985) Plant Mol Biol 4:39–54
Murata N, Miyao M, Hayashida N, Hidaka T, Sugiura M (1988) FEBS Lett 235:283–288
Neuhaus H, Link G (1987) Curr Genet 11:251–257
Neuhaus H (1989) Nucleic Acids Res 17:444
Neuhaus H, Scholz A, Link G (1989) Curr Genet 15:63–70
Neumann EM (1988) Carlsberg Res Commun 53:259–275
Ohyama K, Fukuzuma H, Kohchi T, Shirai H, Sano T, Sano S, Umesoni K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Nature 322:572–574
Oliver RP, Poulson C (1984) Carlsberg Res Commun 49:647–673
Quigley F, Weil JH (1985) Curr Genet 9:495–503
Reiter W-D, Palm P, Yeats S (1989) Nucleic Acids Res 17:1907–1914
Rock CD, Barkan A, Taylor WC (1987) Curr Genet 12:69–77
Shinozaki K, Deno H, Sugita M, Kuramitsu S, Sugiura M (1986a) Mol Gen Genet 202:1–5
Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazowa K, Yamada K, Kusada J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986b) EMBO J 5:2043–2049
Simons RW, Klechner N (1988) Annu Rev Genet 22:567–600
Sugita M, Shinozaki K, Sugiura M (1985) Proc Natl Acad Sci USA 82:3557–3561
Tanaka M, Obokata J, Chunwongse J, Shinozaki K, Sugiura M (1987) Mol Gen Genet 209:427–431
Webber AN, Hird SM, Packman LC, Dyer TA, Gray JC (1989a) Plant Mol Biol 12:141–151
Webber AN, Packman L, Chapman DJ, Barker J, Gray JC (1989b) FEBS Lett 242:259–262
Westhoff P (1985) Mol Gen Genet 201:115–123
Westhoff P, Farchaus JW, Herrmann RG (1986) Curr Genet 11:165–169
Westhoff P, Herrmann RG (1988) Eur J Biochem 171:551–564

Communicated by C. S. Levings III