Published in Biochimica et Biophysica Acta (BBA) 1218, issue 1, 75-81, 1994
which should be used for any reference to this work

Analysis of the 22 kbp long *psb*D-*psb*C gene cluster of *Euglena gracilis* chloroplast DNA: evidence for overlapping transcription units undergoing differential processing

Bernard Orsat [1], Albert Spielmann, Sophie Marc-Martin, Thomas Lemberger [2], Erhard Stutz *

Laboratoire de Biochimie Végétale, Université de Neuchâtel, Chantemerle 18, CH-2000 Neuchâtel, Switzerland

Abstract

The clustered genes *psb*D and *psb*C covering together close to 22 000 nucleotides contain ten and eleven exons, respectively. The corresponding translation products, i.e, Photosystem II core 34 kDa (D2) protein and the CP43 chlorophyll binding protein are highly conserved. Introns vary in lenght from 305 to 4144 nucleotides. The two genes have about 900 nucleotides in common including an intron. To obtain stable mRNAs of about 1400 (*psb*D) and 1500 (*psb*C) nucleotides the pre-transcripts must undergo differential processing and/or splicing events within the overlapping region.

Key words: DNA sequence; Chloroplast; Photosystem II; *psb*D; *psb*C; Transcription; (*E. gracilis*)

1. Introduction

The chloroplast genes *psb*D and *psb*C code, respectively, for the Photosystem II core 34 kDa (D2) and the CP43 chlorophyll binding protein. Both genes were first identified and sequenced in spinach [1,2] and it was shown that the reading frames of the two genes overlap by 50 nucleotides. The N-terminal part of the CP43 protein was sequenced [3] and found to start with *N*-acetyl-*O*-phosphothreonine which corresponds to residue 15 of the sequence predicted by the *psb*C sequence. In *Chlamydomonas reinhardtii* [4] the *psb*C gene has no overlap with *psb*D. It lacks the first 12 amino acids predicted by the spinach *psb*C sequence. The reading frame starts with GUG. This codon is conserved in all *psb*C genes sequenced so far and it was argued that either translation in the spinach gene starts with GUG or the first 14 amino acids of a precursor protein are cleaved to yield a mature protein [3]. In *Euglena gracilis psb*D and *psb*C are also clustered, the two reading frames overlap by 91 nucleotides and the GUG codon equivalent to position 1 (*Chlamydomonas*) and 13 (spinach) is also conserved [5]. However, the anatomical situation is complicated by the presence of introns in the overlapping region.

Transcription of the operon *psb*D-*psb*C was studied in great detail. In barley [6] and tobacco [7], e.g., multiple promoters were observed and transcripts with different 5′ and 3′ ends were identified. Some transcripts start upstream of *psb*D covering both genes while other transcripts start within *psb*D coding only for the CP43 protein, i.e., the dicistronic operon can comprise multiple overlapping transcription units. In *Euglena gracilis* the two clustered genes with multiple introns extend over close to 22 kb [5,8]. Obviously, this brings about unique problems of transcription and transcript processing including splicing.

We show here structural details of the *psb*D-*psb*C operon, define the size of the stable transcripts for both genes, show that the 5′ ends of *psb*C mRNA are within an exon of the *psb*D gene and that the tran-

* Corresponding author. Fax: +41 38 242695.
[1] Present address: Department of Chemistry, M.I.T. Cambridge, MA 02139, USA.
[2] Pressent address: Institut de Biologie Animale, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland.
The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank Data Libraries under the accession number X70810.

script(s) must undergo differential splicing and processing within the overlapping region in order to get functional *psb*D and *psb*C mRNAs.

2. Materials and methods

2.1. Chloroplast DNA and RNA isolation

Growth conditions for *Euglena gracilis* cells (Z-strain) and protocols for chloroplast DNA and RNA purification, and for Southern and Northern blottings were described [9].

2.2. Cloning of an EcoRI-E region

Since we did not achieve to clone the entire *Eco*RI-E fragment (7851 bp) of *Euglena gracilis* chloroplast DNA using both, phage and plasmid vectors, a chromosome walking strategy was used to clone a 5.0 kb DNA region located between a previously mapped *Bgl*II [5] and *Hin*dIII site [10] within *Eco*RI-E. Total *Euglena* DNA was restricted with various restriction enzymes, separated on agarose gel, blotted onto Nylon filter and hybridized to the left part (2.5 kb *Eco*RI-*Bgl*II fragment) or the right part (0.6 kb *Hin*dIII-*Eco*RI fragment) of *Eco*RI-E. Most of the hybridizing fragments could not be cloned, for yet unknown reasons, except a 2.0 kb *Dde*I fragment (hybridizing to the *Eco*RI-*Bgl*II probe) and another 2.9 kb *Dde*I fragment (hybridizing to the *Hin*dIII-*Eco*RI fragment). Both fragments were cloned into pBluescript vectors SK- (Stratagene), creating pEgc24 (left part of *Eco*RI-E) and pEgc25 (right part of *Eco*RI-E). After charaterization by restriction analysis and partial sequencing, total or part of these clones were again used as probes on Southern blots of total *Euglena gracilis* DNA. Again, several fragments hydridized to the probes, amongst a 1.4 kb *Hin*dIII fragment (reacting to a 0.4 kb *Hin*dIII-*Dde*I from pEgc24) and a 2.8 kb *Hin*dIII fragment (reacting to a 1.1 kb *Dde*I-*Asp*700I fragment of pEgc25). These two fragments were cloned into pBluescript SK-(Stratagene), creating pEgc31 and pEgc28, respectively. After characterization by restriction analysis and sequencing, a 0.7 kb *Hin*dIII-*Dde*I fragment from pEgc28 was used as probe to clone a 1.2 kb *Dde*I fragment overlapping the two *Hin*dIII fragments of pEgc31 and pEg28.

2.3. Sequencing strategy and sequence analysis

DNA sequences were determined using the pBluescript Exo/Mung DNA sequencing system following the instructions of the suppliers (Stratagene). DNA sequence analysis was performed using the GCG software package from Wisconsin [11].

2.4. Primer extension analysis

Primer extension studies were performed on total *Euglena gracilis* RNA (10 μg) using the primer EG-4 defined in Fig. 5 and the reverse transcriptase Superscript (Gibco-BRL) following the conditions recommended by the suppliers.

2.5. S1-endonuclease mapping

Approx. 2 μg of *Euglena gracilis* chloroplast RNA were hybridized to 1 μg of ^{32}P-end labelled 762 bp *Hin*cII DNA fragment at 40°C for 16 h. Nuclease S1 was added (100 to 1000 units/ml) and the reaction mix was incubated at 40°C for 1 h. After purification by phenol extraction and ethanol precipitation, an aliquot of the reaction mix was characterized by electrophoresis through a 6% polyacrylamide denaturing gel.

2.6. Reverse transcriptase-polymerase chain reaction

Based on the chloroplast DNA sequence, three synthetic oligonucleotides were synthesized whose sequences are given in Fig. 5. (1) cDNA primers EG-4 and EG-5 complementary to the RNA-like strand in exon 4 of *psb*C and the 3′ untranslated region of *psb*D, respectively, were used to prime cDNA synthesis on total *Euglena gracilis* RNA. (2) A PCR primer (EG-10), complementary to the cDNA-like strand of the presumed overlapping region of *psb*D-*psb*C within exon 10 of *psb*D was used to amplify the cDNAs synthesized from EG-4 or EG-5, respectively. cDNA synthesis was performed with 5 μg of total *Euglena gracilis* RNA with either EG-4 or EG-5 and 2 units of reverse transcriptase (Retrotherm, Epicentre) for 10 min at 40°C, 10 min at 50°C and 10 min at 70°C according to the conditions recommended by the suppliers. One tenth of the synthesized cDNA was directly used for PCR amplification (enzyme Replitherm, Epicentre). Conditions were, 30 cycles (94°C, 30 s; 45°C, 30 s; 72°C, 60 s) in a Perkin-Elmer Cetus DNA thermal cycler. The PCR products were gel-purified, cloned in pBluescript SK+ (Stratagene), and characterized by sequencing.

3. Results and Discussion

In Fig. 1 an overview of the anatomy of the two clustered genes is presented. The genes are situated (5′ → → 3′) between *bch*I (former *ccs*A) with opposite polarity [12] and *trna*L [13]. *psb*D and *psb*C contain ten and eleven exons, respectively, interrupted by introns which vary in size from 305 to 4144 nucleotides (Table 1). The nucleotide sequence coordinates of both genes were published [8]. Three of the large introns

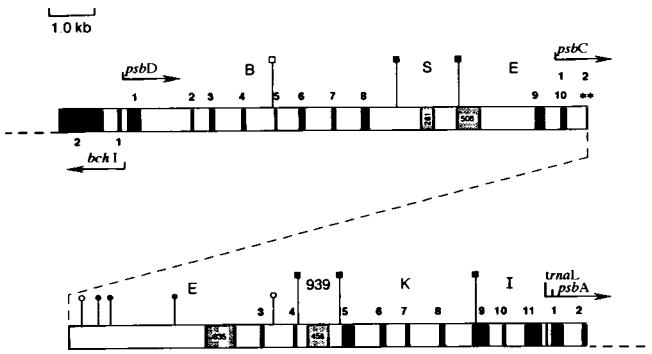

Fig. 1. Mapping and structural features of the clustered genes *psb*D-*psb*C. Black boxes mark exons separated by introns, numbering starts separately for each gene; shaded boxes with number of codons mark ORFs within introns; bent horizontal arrows mark 5′ ends and transcription polarity. New name for *ccs*A is *bch*I [8]; **, two stop codons at the 5′ end of the tail (*psb*D); ■ = *Eco*RI sites separating previously mapped fragments [23]; □ = *Sal*I; restriction sites mentioned in the sequencing protocol: ○ = *Hin*dIII; ● = *Dde*I.

contain major ORFs of yet unknown function. It may be noteworthy that ORF506 (*psb*D, intron 8) is similar in a 452 amino acid overlap with ORF608 (*pet*D, intron 1) of a green alga (genus, *Ankistrodesmus*) [14]. Both ORFs contain Zn-finger motives. Orf608 shows in addition significant homology with reverse transcriptases [14] what is not the case for ORF506.

From Northern analyses using different DNA probes it became evident that the stable mRNAs of *psb*D [5] and *psb*C [10,15] are in the range of 1.4 and 1.5 kb, respectively. Both genes are transcribed in the dark (not shown) and transcription patterns are very complex. We show and compare in Fig. 2 Northern imprints obtained by hybridizing purified chloroplast RNA from light grown *Euglena* cells with *psb*D and *psb*C DNA probes. The difference in the two patterns is remarkable in at least two aspects: (1) the *psb*D pattern is more complex with a larger number of RNA intermediates than the *psb*C pattern and (2) the largest

Table 1
Structural features of the genes *psb*D and *psb*C

*psb*D			*psb*C		
No.	exon (n)	Intron (n)	No.	exon (n)	Intron (n)
1	243	1098	1	14*	*543*
2	35	364	2	11	**4144**
3	84	605	3	48	671
4	74	651	4	45	**1605**
5	41	498	5	240	590
6	115	606	6	107	448
7	63	580	7	70	668
8	157	**3658**	8	98	621
9	118	373	9	348	305
10	126	*543*	10	52	423
11	≈ 270 (tail)		11	350	

n, nucleotides; *, provided GUG is the start codon; bold introns contain ORFs; common intron is in italics.

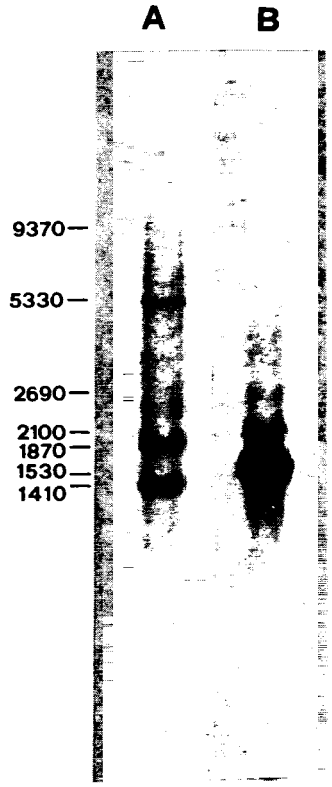

Fig. 2. Northern blot analysis of *psb*D and *psb*C transcripts. Purified chloroplast RNA was hybridized to a *Sal*I-*Eco*RI fragment (2600 bp) covering the central part of *psb*D (including exons 5, 6, 7, 8) (pattern A) and to a *Eco*RI-*Eco*RI fragment (3551 bp) covering *psb*C exons 5, 6, 7, 8 and parts of 9 (pattern B). Fragment length (nucleotides) calculated from markers is given on the margins.

*psb*D pre-RNA is in the range of 9.3 kb while the largest *psb*C pre-RNA in this and other preparations [10,15] is in the range of 3 kb. Since the results were obtained under strictly comparable conditions and since both genes have about the same size and number of introns their transcription and in particular splicing dynamics must be differentially controlled.

The 5′ end of the *psb*D mRNA was determined by S1-nuclease protection analysis using a *Hinc*II-*Hinc*II (762 bp) DNA fragment which covers 145 nucleotides of the coding part downstream of the AUG start codon. According to the size of the protected fragment seen in Fig. 3 the 5′ end is 79 nucleotides upstream of the start codon. The same result was obtained with primer extension experiments (not shown). The integrated length of *psb*D exons is 1056 nucleotides. Considering the length of the stable mRNA and of the untranslated leader we calculate the untranslated tail to be about 270 nucleotides (see also below).

We show in Fig. 4 the sequence of the DNA region with the identified *psb*D transcription start site. We tentatively mark possible −10 and −35 promoter elements, well knowing that so far no *Euglena* chloroplast promoter sequences have been adequately charac-

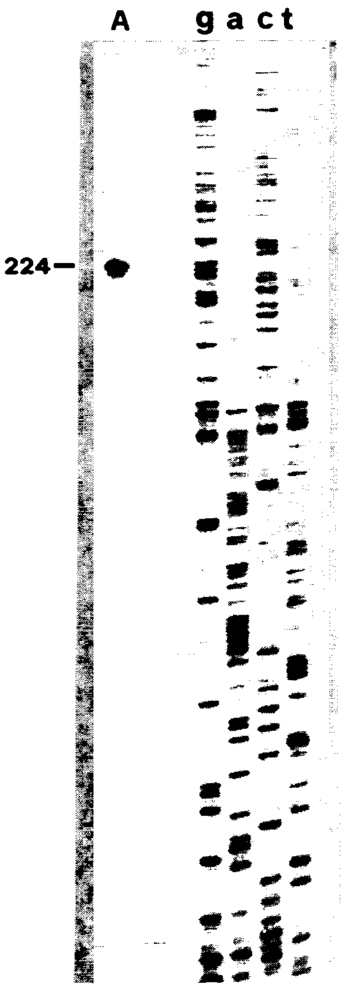

Fig. 3. S1-endonuclease protection analysis of *psb*D mRNA. A *Hin*cII-*Hin*cII double strand DNA fragment (762 nucleotides) covering 145 nucleotides of the 5′ part of *psb*D exon 1 was hybridized to purified chloroplast RNA. Size of the protected fragment in lane A was calculated from the sequence ladder of a cloned DNA fragment.

```
12961 ttgcatattagtaccttaatgaaaagtagtatgtatttgaaagcaagtactatttggtaa
13021 gaggtatttttattaaataccgacttctacTTTGAAACTTTTTATACTAAAAATATTCTT
                     psbD [exon 10] F  E  T  F  Y  T  K  N  I  L
13081 TTAAATGAAGGTATTAGAGCTTGGATGGCTGCACAAGATCAGCCTCATGAACAATTTATA
      L  N  E  G  I  R  A  W  M  A  A  Q  D  Q  P  H  E  Q  F  I
         m  k  v  l  e  l  g  w  l  h  k  I  s  l  m  n  n  l  y
13141 TTCCCGGAGGAGGTTCTTCCACGTGGAAACGCTCTTgtttgtttttgtatctttttaata
      F  P  E  E  V  L  P  R  G  N  A  L
       s  r  r  r  f  f  h  V  E  T  L  F  [exon 1] psbC
13201 gtcgaaaatactacaaaaaatttttatttattaggtttttttcctattgaaatttatacta
13261 atttttttgtaaaagcagttgtaagatctgtgaaaaggagattatgaagttttataaaat
13321 actttaaactattattattattgtagtgtatttttcttttatttttttagtttgttaaaac
13381 ttgctaatttaaaaaagagcttcctttttgctgtaatattattttttattctgaaaatt
13441 taataatgaattatttttgtctatgctattaaaaattatgaataatctaataattggatt
13501 tttaaaacgctatttttgttaatatgtgacaaaaaaatgtcaaaactggtttttatttgt
13561 tagttttgtttttgttttaaaaaattaattgtttgaaagaattatggtttcaaaaggcct
13621 caaaaatttaaaatcgataaaatatagcttaataagcaaaaaagcttctgtaaagttagg
13681 aggaaatttgattattttttctaatcttttaatcctatTAATAAAAAATgtgtggcatg
          psbD [tail with 5'end stop codons] *  *
                             psbC [exon 2]    M  K  K  L
13741 gattccaataatcgctaggtttaggattaaaaaataatttttaacttaagtcatt.....

17822 tgcaatgaattatttgtgcaagtatagtttggaaaacggattttcgatttaacTAACTGT
                                                psbC [exon 3] T  V
17882 AGGTGGTCGTGATCAGGAATCAACTGGATTTGCTTGGTGGGgtgcgaaacgtttt.....
        G  G  R  D  Q  E  S  T  G  F  A  W  W  A
18542 cgtaaagtgttttggagggaaactttagaattttaatttttctatcttaatCAGGTAAT
                                            psbC [exon 4]   G  N
18602 GCTAGACTTATAAATGTATCAGGTAAACTTTTGGGTGgtgtgttttgaattctaa.....
        A  R  L  I  N  V  S  G  K  L  L  G  A
```

Fig. 5. DNA sequences of the *psb*D-*psb*C translation overlap (frameshift) and of *psb*C regions carrying exons 3 and 4. Small and capital letters represent introns and exons, respectively; consensus 5′ ends of introns are in italics; primers EG10, EG5, and EG4 (top to bottom) used in PCR are doubly underlined; → → mark two *psb*C 5′ ends; GTG (bold) marks the most likely translation start site of CP43; capitalized amino acid sequences represent exons, small letters represent in frame translation of *psb*C upstream of GTG; numbers on the margins correspond to registered sequence coordinates (see legend to Fig. 4).

```
tttattcatagtaaattgatagattgttatgcaatatttttgcttgtgttccctttttt
aaataaGTAtcatttaactatctaacaatacgttataaaaaacgaacacaagggaaaaaa
 K  N  M                                       ↔ bchI
        psbD ↔
agtatataaatagggggatcttaatacctaaaattagctctttagttaatattagagatta
tcatatatttatcccctagaattatggattttaatcgagaaatcaattataatctctaat

cttattttatattacaaaatattatttATGacttttacagatttaaatactgaaaataaa
gaataaaatataatgttttataataaatactgaaaatgtctaaatttatgacttttattt
                              M  T  F  T  D  L  N  T  E  N  K
```

Fig. 4. DNA sequence of a segment upstream of *psb*D. The 5′ end is marked by → → and the mRNA-like strand is in italics. Promoter consensus elements are overlined. The start codon is in capital letters. For convenience we mark the 5′ end of the *bch*I mRNA with opposite polarity (Stauffer and Stutz, unpublished). The displayed sequences correspond to nucleotide positions 3541 to 3720 of the complete *Euglena gracilis* chloroplast DNA sequence submitted to the EMBL/GenBank Data Libraries, Heidelberg, accession No. X70810.

terized. The first 20 positions upstream of the translation start codon are very rich in AT (95%) lacking a canonical ribosome binding site. Such AT rich sequences are typical for *Euglena* chloroplast mRNAs and may very well function in ribosome binding as was recently discussed [16].

A unique structural feature concerns the 3′ end of *psb*D and the 5′ end of *psb*C (Fig. 5). They share a common intron of 543 nucleotides. Intron 10 (*psb*D) cuts the C-terminal exon 10 from exon 11 which carries at its 5′ end two stop codons followed by a non coding sequence (tail) of about 270 nucleotides (consult also Fig. 8). Intron 1 (*psb*C) separates the N-terminal exon 1 with GUG as most likely start codon from a very short exon 2 of eleven nucleotides which is 4144 nucleotides away from the next *psb*C exon 3. This intron (twintron, composite structure, Hallick and Stutz, un-

published observation) is the largest chloroplast DNA intron yet detected in a protein coding gene.

Experimental evidence for joining (splicing) the tail to exon 10 of *psb*D and for joining exons 1, 2, 3, 4 of *psb*C is given in Fig. 6. It shows the sequences of the reverse transcriptase PCR products having used as starting primers the three oligonucleotides, EG – 4 within exon 4 of *psb*C, EG – 5 within the untranslated tail of *psb*D and EG – 10 within a region common to both transcripts (consult also Figs. 5 and 8). The sequences displayed in Fig. 6 match, respectively, the spliced mRNA of *psb*D and of *psb*C (junction sites of exons 1, 2, 3, 4).

The 5′ end of the *psb*C mRNA was determined by primer extension experiments (Fig. 7). The position of the strong band corresponds to a *psb*C mRNA with a 5′ end located within exon 10 of *psb*D, 41 nucleotides upstream of the presumed GUG start codon (see Fig. 5). An additional faint band is seen corresponding to a mRNA 5′ end 81 positions upstream of GUG. This second site could be the transcription start site. If so, the primary transcript of *psb*C is shortened by about 40 nucleotides at its 5′ end. Interestingly enough the latter position matches a processed 5′ end of tobacco *psb*C mRNA [7].

The question remains whether the clustered *Euglena* genes are cotranscribed as seen, e.g., in spinach [1,2]. In Northern prints we could never detect any

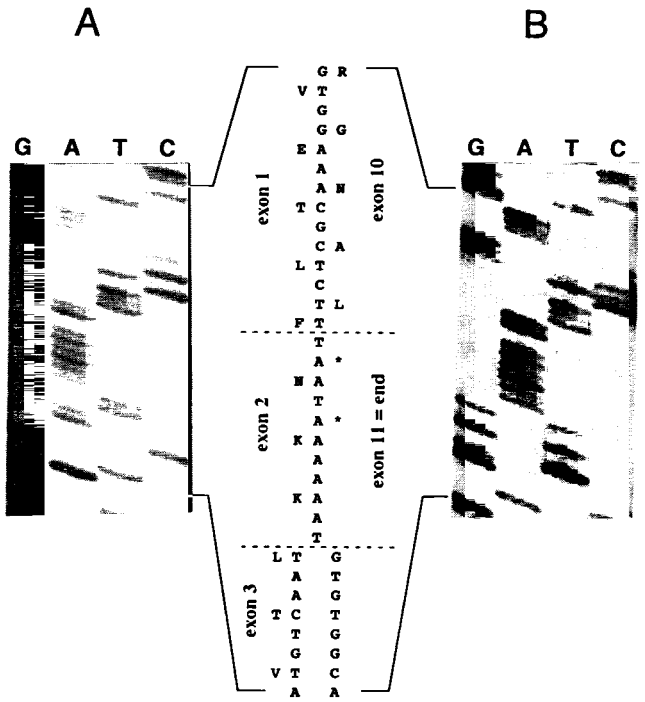

Fig. 6. DNA sequences of reverse transcriptase-polymerase chain reaction products. (A) Sequence of the amplified reverse transcript EG4-EG10 (*psb*C); (B) Sequence of the amplified reverse transcript EG5-EG10 (*psb*D); Nucleotide and amino acid sequences are indicated including exon boundaries.

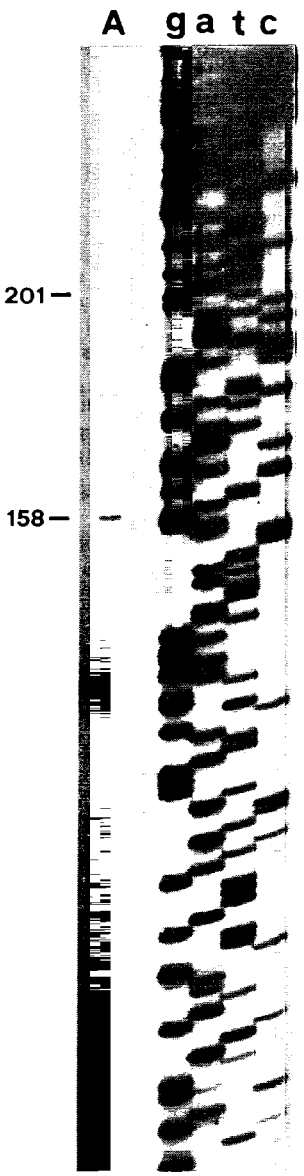

Fig. 7. Primer extension analysis of *psb*C mRNA. A major (158) and a minor (201) reverse transcription product obtained using the primer EG4 (see Fig. 5) are shown in lane A. Their size is calculated from the sequence ladder of a known clone.

precursor transcripts larger than the ones shown in Fig. 2, nor was a *psb*D-*psb*C cotranscript ever seen in extensive electron microscopic studies [17]. On the other hand, eleven introns were clearly seen in the EM as loops within *psb*D what is in perfect line with our sequence data. Some of the smaller but not the largest intron of *psb*C were also identified. According to Dr. B. Koller, University of Lund, it is unlikely that under the required denaturing-renaturing conditions the loop corresponding to the large intron 2 is detected since the very short exon 2 would not form a stable hybrid (personal communication). These and our data suggest that *psb*C transcription starts within *psb*D. However, a primary *psb*D-*psb*C cotranscript of at least 22 kb un-

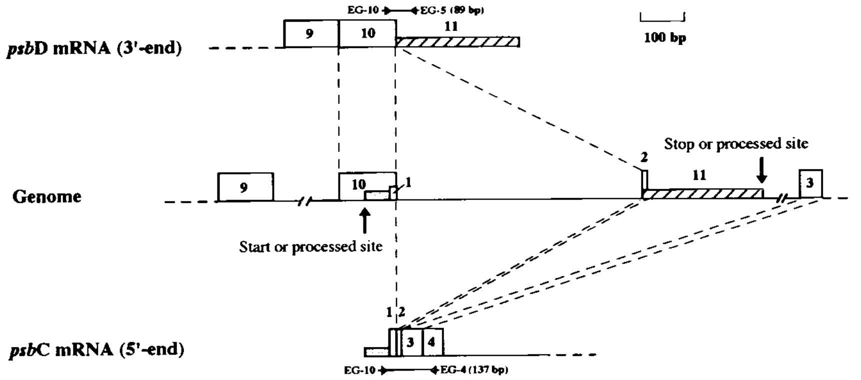

Fig. 8. Overview of pre-mRNA processing and/or splicing events in the overlapping region and its vicinity leading to mature *psb*D and *psb*C mRNAs. Open boxes represent exons numbered as shown in Table 1; stippled box is untranslated leader (*psb*C); hatched box is the tail with stop codon (*psb*D); EG10-EG4 and EG10-EG5 give the position of the amplified cDNAs.

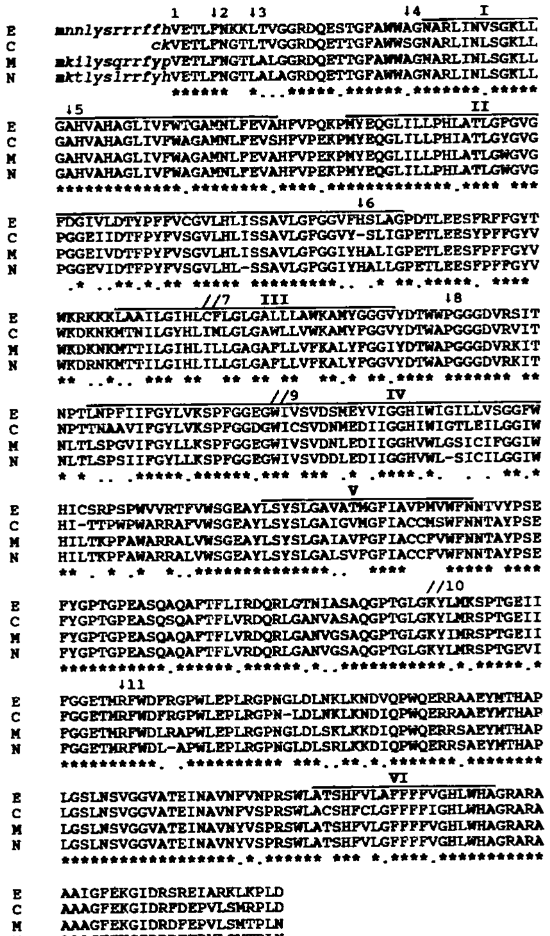

Fig. 9. Sequence alignment of *Euglena gracilis* CP43 protein with chloroplast counterparts. E, *Euglena gracilis*; C, *Chlamydomonas reinhardtii* [4]; M, *Marchantia polymorpha* [23]; N, *Nicotiana tabacum* [24]; the sequences are aligned (CLUSTAL) starting with GUG (V) which can code for M_i. In frame upstream regions are given in italics; six hydrophobic domains are overlined (roman numbers); *Euglena* exons are numbered 1 to 11; ↓ and || mark intron insertion sites with flush and split codons, respectively; * and · indicate perfect and conserved matches, respectively.

dergoing rapid processing cannot be excluded (see below).

The two genes have at least 900 nucleotides in common whether an integral primary cotranscript exists or not (Fig. 5). In Fig. 8 possible processing and splicing scenarios of the transcripts in the overlapping region are summarized. (a) Transcription stops within intron 2 downstream of the *psb*D translation stop codon. The 5′ splice site of intron 2 becomes inactive since the transcript does not have the secondary structure information required for correct splicing of group II introns [18,19]. But on the other hand the presumed transcription stop signal has to be ignored in order to obtain the *psb*C mRNA. (b) Intron 2 is entirely transcribed but an endonucleolytic activity yielding the 3′ end of *psb*D mRNA competes with the multistep splicing event (twintron) required to join exons 2 and 3 of *psb*C. (c) Similarly, a common pre-transcript is cleaved to yield the 5′ end of *psb*C or exon 10 is spliced to exon 9 (*psb*D). (d) *psb*C transcription starts upstream of the presumed start codon GUG and a possible pre-transcript undergoes 5′ exonucleolytic processing. We have not yet made in vitro capping experiments to clear that question. In any case whatever scenario turns out to be true (also more than one pathway is possible, see Refs. 6 and 7) the pre-transcripts are differentially tailored within the overlapping region. This may require specific controlling elements or it may be more of a stochastic process [20].

We previously reported that the *Euglena gracilis* D2 protein (252 residues) is highly conserved [5]. Also the CP43 protein (461 residues) is identical to about 80% with chloroplast counterparts (Fig. 9). It is reasonable to assume that the start codon is GUG following the arguments forwarded in the introduction. Furthermore, the 5′ end of the most prominent mRNA is only seven nucleotides upstream of the AUG start codon leaving no place for ribosome binding. On the other

hand upstream of the GUG codon there is a well preserved ribosome binding site as was already noticed, e.g., in spinach [1,2] and in the unicellular red alga *Cyanidium caldarium* [21].

In contrast to all other chloroplast *psb*C genes the *Euglena* gene is interrupted by ten introns some of which are extremely large with important ORFs. Chloroplast genes evolved before the divergence between eubacteria and eukaryotes, i.e., the numerous *Euglena* introns are most likely not remnants of gene shuffling ('introns early') [22]. Rather we believe [8] that *Euglena* introns are descendants of mobile elements. A test in point may be the position of the introns within *psb*C. The CP43 protein contains six hydrophobic *trans*-membrane domains which are crucial for maintaining correct spacing within the thylakoid membrane [1]. As shown in Fig. 9 three out of six domains are cut by introns what one would not expect if introns were ancient marks of gene rearrangements.

Acknowledgments

We are grateful to Dr. P.E. Montandon who was involved in the early phase of sequencing the DNA fragment *Eco*RI-E. This project received support from the Swiss National Science Foundation (to E.S.) and the Roche Foundation, Basel.

References

[1] Holschuh, K., Bottomley, W. and Whitfeld, P.R. (1984) Nucleic Acids Res. 12, 8819–8834.
[2] Alt, J., Morris, J., Westhoff, P. and Herrmann, R.G. (1984) Curr. Genet. 8, 597–606.
[3] Michel, H., Hunt, D.F., Shabanowitz, J. and Bennett, J. (1988) J. Biol. Chem. 263, 1123–1130.
[4] Rochaix, J.D., Kuchka, M., Mayfield, S., Schirmer-Rahire, M., Girard-Bascou, J. and Bennoun, P. (1989) EMBO J. 8, 1013–1021.
[5] Orsat, B., Chatellard, Ph. and Stutz, E. (1992) in Research in Photosynthesis (Murata, N., ed.), Vol. III, pp. 255–258, Kluwer Academic Publishers, Dordrecht.
[6] Berends Sexton, T., Christopher, D.A. and Mullet, J.E. (1990) EMBO J. 9, 4485–4494.
[7] Yao, W.B., Meng, B.Y., Tanaka, M. and Sugiura, M. (1989) Nucleic Acids Res. 17, 9583–9591.
[8] Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. (1993) Nucleic Acids Res. 21, 3537–3544.
[9] Montandon, P.E., Knuchel-Aegerter, C. and Stutz, E. (1987) Nucleic Acids Res. 15, 7809–7822.
[10] Montandon, P.E., Vasserot, A. and Stutz, E. (1986) Curr. Genet. 11, 35–39.
[11] Devreux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[12] Orsat, B., Monfort, A., Chatellard, P. and Stutz, E. (1992) FEBS Lett. 303, 181–184.
[13] Keller, M. and Stutz, E. (1984) FEBS Lett. 175, 173–177.
[14] Kück, U. (1989) Mol. Gen. Genet. 218, 257–265.
[15] Hollingsworth, M.J., Johanningmeier, U., Karabin, G.D., Stiegler, G.L. and Hallick, R.B. (1984) Nucleic Acids Res. 12, 2001–2017.
[16] Wang, C-C., Roney, W., Alston, R.L. and Spremulli, L. (1989) Nucleic Acids Res. 17, 9735–974.
[17] Koller, B. and Delius, H. (1984) Cell 36, 613–622.
[18] Michel, F., Umesono, K. and Ozeki, H. (1989) Gene 82, 5–30.
[19] Copertino, D.W. and Hallick, R.B. (1991) EMBO J. 10, 433–442.
[20] Westhoff, P. and Herrmann, R.G. (1988) Eur. J. Biochem. 171, 551–564.
[21] Maid, U. and Zetsche, K. (1992) Plant Mol. Biol. 19, 1001–1010.
[22] Dorit, R.L., Schoenbach, L. and Gilbert, W. (1990) Science 250, 1377–1382.
[23] Hallick, R.B. and Buetow, D.E. (1989) in The Biology of *Euglena* (Buetow, D.E., ed.), Vol. IV, pp. 351–414, Academic Press, San Diego.
[24] Umesono, K., Inokuchi, H., Ohyama, K. and Ozeki, H. (1984) Nucleic Acids Res. 12, 9551–9565.
[25] Sugiura, M. (1987) Bot. Mag. Tokyo 100, 407–436.