The EMBO Journal vol.3 no.12 pp.2753-2762, 1984

Chlamydomonas reinhardii gene for the 32 000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat

Jeanne M.Erickson, Michèle Rahire and Jean-David Rochaix

Departments of Molecular Biology and Plant Biology, University of Geneva, 30, quai Ernest-Ansermet, 1211 Geneva 4, Switzerland

Communicated by J.-D.Rochaix

The chloroplast *psb*A gene from the green unicellular alga *Chlamydomonas reinhardii* has been localized, cloned and sequenced. This gene codes for the rapidly-labeled 32-kd protein of photosystem II, also identified as as herbicide-binding protein. Unlike *psb*A in higher plants which is found in the large single copy region of the chloroplast genome and is uninterrupted, *psb*A in *C. reinhardii* is located entirely within the inverted repeat, hence present in two identical copies per circular chloroplast genome, and contains four large introns. These introns range from 1.1 to 1.8 kb in size and fall into the category of Group I introns. Two of the introns contain open reading frames which are in-frame with the preceding exon sequences. We present the nucleotide sequence for the *C. reinhardii psb*A 5'- and 3'-flanking sequences, the coding region contained in five exons and the deduced amino acid sequence. The algal gene codes for a protein of 352 amino acid residues which is 95% homologous, excluding the last eight amino acid residues, with the higher plant protein.

Key words: group I introns/herbicide resistance/nucleotide sequence/photosynthesis/RNA maturases

Introduction

One of the most abundantly synthesized proteins in the chloroplast is a thylakoid membrane polypeptide of ~32-35 kd (Eaglesham and Ellis, 1974; Bottomley *et al.*, 1974). This protein, generally referred to as the rapidly-labeled 32-kd protein of photosystem II (PSII), has been extensively characterized in many higher plant species and in algae (Hoffman-Falk *et al.*, 1982; Delepelaire, 1983, 1984). It appears to be synthesized as a precursor of ~34 kd (Grebanier *et al.*, 1978; Edelman and Reisfeld, 1980; Reisfeld *et al.*, 1982) which is then processed at the carboxyl terminus (Marder *et al.*, 1984) to produce a polypeptide of ~32 kd. However, analysis is complicated by the fact that this polypeptide migrates in acrylamide gel systems as a diffuse band that is poorly stained with Coomassie brilliant blue, and is one of several thylakoid membrane proteins with a mol. wt. in this range (Bottomley *et al.*, 1974; Chua and Gillham, 1977; Kuwabara and Murata, 1979; Vermaas *et al.*, 1984). The rapid rate of synthesis of this protein is light induced (Bedbrook *et al.*, 1978; Mattoo *et al.*, 1984). However, a high rate of turnover in the light (Mattoo *et al.*, 1984; Wettern *et al.*, 1983) prevents the accumulation of protein. Thus, adequate quantities of purified protein have been difficult to obtain for sequencing or for antibody production.

The exact function of the 32-kd protein, which can be found in isolated PSII particles, is still being investigated. Because of its proposed role as part of a protein-quinone complex (Arntzen *et al.*, 1982) that acts as the secondary stable electron acceptor of PSII Velthuys and Amesz, 1974), the 32-kd protein has recently been called the Q_B protein. The rapid turnover of the 32-kd protein may in fact be a consequence of its normal function in accepting electrons (Mattoo *et al.*, 1981; Ohad *et al.*, 1984). This protein does play a central role in determining resistance to certain herbicides, such as atrazine (2-Cl-4-ethylamino-6-isopropylamino-*s*-triazine) and diuron of DCMU[3-(3,4-dichlorophenyl)-1,1-dimethylurea] (Trebst, 1980), which block PSII electron transport: three different herbicide-resistant mutants of *Chlamydomonas reinhardii* (Erickson *et al.*, 1984, and unpublished data) and herbicide-resistant biotypes of *Amaranthus hybridus* (Hirschberg and McIntosh, 1983) and *Solanum nigrum* (Hirschberg *et al.*, 1984) all have an altered 32-kd protein, as deduced from DNA sequence analysis of the gene. This alteration greatly reduces the level of normal binding of azido[^{14}C]atrazine (Steinback *et al.*, 1981) to the 32-kd protein (Pfister *et al.*, 1981; Erickson *et al.*, 1984).

The gene which codes for the 32-kd protein, called *psb*A, is located in the chloroplast genome of higher plants and algae. *psb*A has been sequenced in spinach and *Nicotiana debneyi* (Zurawski *et al.*, 1982), pigweed (Hirschberg and McIntosh, 1983), maize (McIntosh, personal communication), soybean (Spielmann and Stutz, 1983), mustard (Link and Langridge, 1984), *N. tabacum* (Suqita and Sugiura, 1984) and *Euglena* (Karabin *et al.*, 1984; Keller and Stutz, 1984). The deduced protein sequences are highly conserved between species, with no differences in the spinach and *Nicotiana* proteins, and only one, two or three amino acid changes in pigweed, soybean and mustard, respectively. The *Euglena* protein is 87% homologous to that of higher plants. The gene for a protein having ~88% homology with the *psb*A gene product has been sequenced from the cyanobacteria *Anabaena* (Curtis and Haselkorn, 1984) and *Fremyella* (Mulligan *et al.*, 1984).

Here we report the isolation and sequence analysis of *psb*A from the green, unicellular alga *C. reinhardii*. Although this gene codes for a protein that is 93% homologous overall with that of higher plants, the structure of the *C. reinhardii* gene is remarkably different. The higher plant gene is located in the large single copy region of the chloroplast genome and is uninterrupted, while the algal gene is located in the inverted repeat, thus present in two copies per genome, and each algal gene contains four introns.

Results

psbA location and organization in C. reinhardii

Previous studies using hybrid-arrest translation had shown that the message for a thylakoid membrane protein of 32-35 kd hybridized to the chloroplast *Eco*RI restriction fragment R14 (Malnoe *et al.*, 1979). To determine whether this protein was, in fact, comparable with the 32-kd protein of PSII identified in higher plants, and to identify the *C. reinhardii* gene coding for this protein, cloned R14 DNA was hybridized with a nick-translated DNA fragment (described

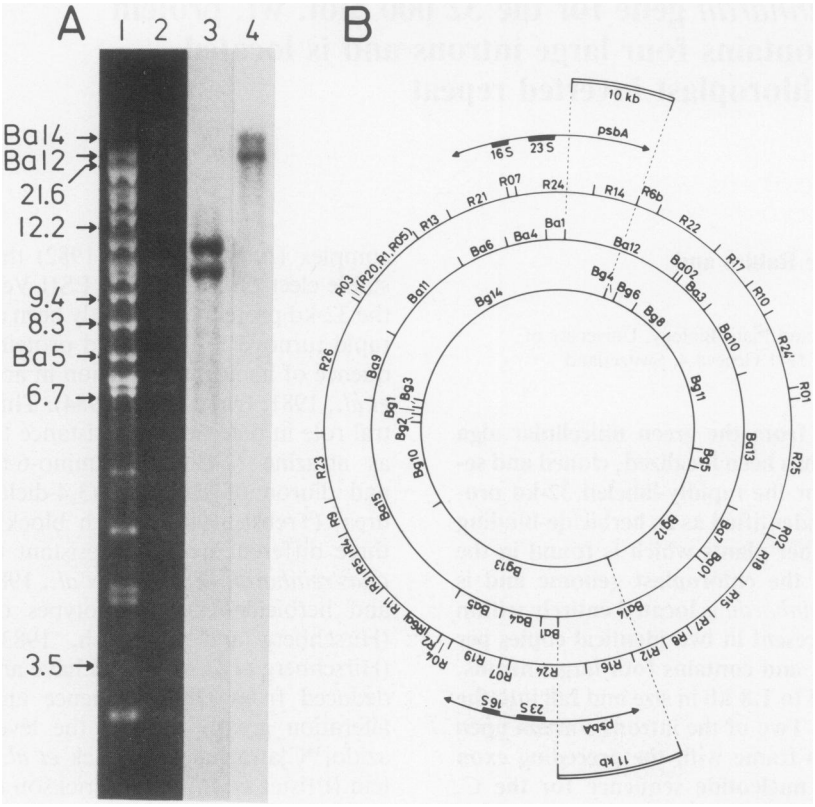

Fig. 1. (a) *C. reinhardii* chloroplast DNA digested with *Bam*HI/*Bgl*II **(lanes 1,3)**, or with *Bam*HI **(lanes 2,4)** was electrophoresed on a 0.6% agarose gel and stained with ethidium bromide **(lanes 1,2)**. The autoradiogram in **lanes 3** and **4** represents a Southern transfer of **lanes 1** and **2** after hybridization with a mixture of two [^{32}P]-nick-translated fragments internal to *Eco*RI restriction fragment R14 (fragments: 1.8-kb *Xba*I, 1.3-kb *Xba*I/*Eco*RI, see Figure 2b). Migration of lambda DNA size markers is indicated with arrows at the left margin, as is the position of chloroplast *Bam*HI restriction fragments Ba14, Ba12 and Ba5. Similar hybridization results were obtained when a 1.2-kb *Sma*I/*Xba*I fragment from pSOCS15, containing coding sequences internal to the spinach *psb*A gene and 220 bp of 5′-flanking sequence (Zurawski *et al.*, 1982), was used as a probe. **(B)** Circular map of the *C. reinhardii* chloroplast genome (Rochaix, 1978), indicating the *Bam*HI/*Bgl*II fragments of ~10 and 11 kb that contain *psb*A and span one end of the inverted repeat. Restriction fragments for *Eco*RI (R), *Bam*HI (Ba) and *Bgl*II (Bg) are indicated in the three concentric circles. The 16S and 23S rRNA genes are indicated and the approximate extent of the inverted repeat is marked with arrows.

in Figure 1 legend) from pSOCS15 which contained the spinach *psb*A gene (Zurawski *et al.*, 1982). In addition, total *C. reinhardii* chloroplast DNA was digested with *Eco*RI, with *Bam*HI or with *Bam*HI/*Bgl*II, electrophoresed on agarose gels, transferred to nitrocellulose and hybridized to the same spinach *psb*A probe. Results of these heterologous hybridizations confirmed that R14 contained sequences highly homologous to the spinach *psb*A. However, both the spinach probe and an internal probe from R14 hybridized not only to restriction fragment R14, but also to restriction fragments R16, R24, Ba12, Ba14 and *Bam*HI/*Bgl*II fragments of 10 and 11 kb (see Figure 1). Thus, there appeared to be two copies of *psb*A in the *C. reinhardii* chloroplast genome. Using a 475-bp *Sma*I-*Pst*I fragment from the spinach *psb*A gene (coding for the first 85 amino acid residues of the protein) as a 5′ probe and a 450-bp *Hae*III fragment which codes for the last 75 residues as a 3′ probe (Zurawski *et al.*, 1982), the 5′ and 3′ ends of the *C. reinhardii* gene were found to be ~7 kb apart. The 3′ end of the algal gene is located in R24, a restriction fragment contained entirely within the inverted repeat. To ensure that a contiguous copy of the algal gene could be analyzed, chloroplast fragments extending from within R24 to beyond the bounds of the inverted repeat were cloned. Though it is somewhat problematical, if not impossible, to distinguish between sequences contained on both halves of the inverted repeat (see Discussion), for the sake of descriptive clarity we will refer to the cloned *psb*A contained in Ba14 or the 11-kb *Bam*HI/*Bgl*II fragment as copy 1, and cloned *psb*A found in Ba12 or the 10-kb *Bam*HI/*Bgl*II fragment as copy 2 (Figure 1a). Thus, *psb*A copy 1 was initially cloned from *C. reinhardii* strain *cw*15 into the lambda vector λEMBL4 (see Materials and methods) as a *Bam*HI/*Bgl*II fragment of 11 kb. Copy 2 was isolated from a recombinant plasmid containing *cw*15 chloroplast restriction fragment Ba12 (Rochaix, 1978). To facilitate sequencing, appropriate subclones of both gene copies were constructed in plasmid vectors as described in Materials and methods.

The restriction map of the inverted repeat region, giving the location of the rDNA genes and *psb*A, is shown in Figure 2a. *psb*A contains five exons, interrupted by four introns of ~1.35, 1.4, 1.1 and 1.8 kb. This gene is transcribed towards the ribosomal DNA, and appears to be entirely located within the inverted repeat. Fine restriction mapping and DNA sequence analysis (unpublished data) show that the end of the inverted repeat is ~1 kb 5′ to the first ATG in the open reading frame of *psb*A. Plasmid and lambda recombinant clones containing either *psb*A copy 1 or copy 2 can be readily distinguished from each other by digesting cloned DNA with *Hind*III and *Eco*RI, which produces a 1.4-kb *Hind*III/*Eco*RI fragment in copy 1 (R16), and a 1.1-kb *Hind*III fragment in copy 2 (R14), as indicated in the divergent map at the left of Figure 2a.

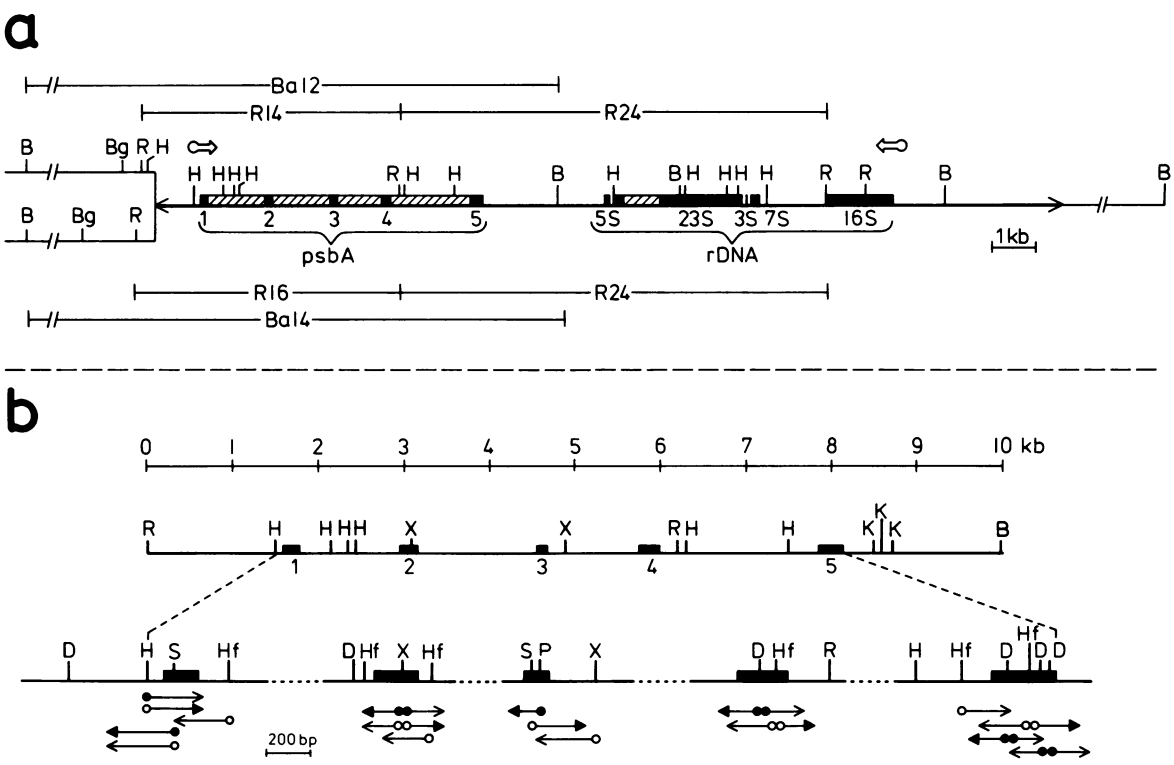

Fig. 2. **(a)** Restriction map of the chloroplast inverted repeat (marked with thick line, flanked by dark arrows) showing the 16S, 7S, 3S, 23S and 5S rRNA genes (▬), the five exon regions of *psb*A (▬), the introns in *psb*A and the 23S rRNA gene (▨), and the directions of transcription (⇨) for the rRNA cistrons and *psb*A. Location of *Eco*RI restriction fragments R24, R16 and R14, and *Bam*HI fragments Ba12 and Ba14 are indicated above and below (see also Figure 1A). The difference between R16 and R14, Ba14 and Ba12, is shown by the divergence at the far left of the linear map. Restriction sites for *Bam*HI (B), *Bgl*II (Bg), *Eco*RI (R) and *Hind*III (H) are as indicated. Size bar represents 1 kb. **(b)** Sequencing strategy for the five exon regions of *psb*A. Fragments eluted from acrylamide gels and digested with the appropriate enzymes were labeled at the 3′ (○→) or 5′ (●→) end, and strand separated before sequencing *via* the Maxam-Gilbert chemical cleavage reactions (Maxam and Gilbert, 1980). In general, both strands of each exon were sequenced. Exceptions are the first 42 bp of exon 3, and the 51 bp of exon 5 between two *Dde*I sites. In both cases, the one strand was sequenced several times with no apparent ambiguities in sequence. Fragments used in S1 protection mapping are indicated with dark arrow heads (—▶). Restriction sites for *Eco*RI (R), *Hind*III (H), *Xba*I (X), *Kpn*I (K), *Bam*HI (B), *Dde*I (D), *Sau*3A (S), *Hinf*I (Hf) and *Pvu*II (P) are as indicated. Size scale for the continuous map is given in kb. Size bar for the lower, expanded map is 200 bp.

Nucleotide sequence and deduced protein sequence

A sequencing strategy for *psb*A was developed by end-label partial-digestion mapping (Smith and Birnstiel, 1976). The five exon regions of *psb*A were sequenced (Maxam and Gilbert, 1980) by the strategy shown in Figure 2b. Exons 1–4 were sequenced for copy 1 of *psb*A (contained in R16) from *C. reinhardii cw*15, and for copy 2 of *psb*A (contained in R14) from DCMU4, a herbicide-resistant *C. reinhardii* mutant (Erickson *et al.*, 1984a). Exon 5 was sequenced for copy 1 (11-kb *Bam*HI/*Bgl*II fragment, see Figure 1) and copy 2 (Ba12) from both *cw*15 and DCMU4. No differences were found between the two *psb*A copies in the first four exons. For the fifth exon, both copies were identical within a given strain. The only difference between strains was a single base substitution, resulting in a deduced amino acid change of serine in *cw*15 to alanine in the herbicide-resistant mutant (Erickson *et al.*, 1984a). It appears, therefore, that within a given strain, the exon sequences of copy 1 and copy 2 are identical.

The nucleotide sequences of the exon coding regions of *C. reinhardii psb*A are given in Figure 3, along with the deduced amino acid sequence beginning with the first methionine in the open reading frame. The protein sequence is compared with that of spinach and *N. debnyi*. The serine residue at position 264 is the residue that is altered to alanine in DCMU4. The algal gene codes for a protein of 352 amino acids, as compared with 353 for the higher plant gene.

Codon usage in the *C. reinhardii psb*A, as shown in Table I, involves only 36 of the 61 possible codons. Codons ending in T, A or C each comprise between 28% and 35% of the 352 codons, while codons ending in G account for 8% of the total. This strong bias against G in the third position is seen in *C. reinhardii psb*D (Table I, in parentheses), which codes for another PSII thylakoid membrane polypeptide (Rochaix *et al.*, 1984). The codon usage for these two algal chloroplast genes are quite similar except for isoleucine and valine (see Table I). All *psb*A genes sequenced to date (see Introduction) terminate with TAA.

The 5′ and 3′ ends of the *psb*A mRNA were determined by S1 mapping (Materials and methods, fragments shown in Figure 2b) and are indicated on the DNA sequence in Figure 3. Two putative transcription initiation sites are seen, one at −34 bp and the other at −90 bp upstream from the first ATG of the open reading frame (Figure 4). This phenomenon was seen in the *A. hybridus psb*A gene (Hirschberg and McIntosh, 1983), as well as several other chloroplast genes (Dron *et al.*, 1982; Rochaix *et al.*, 1984). S1 mapping of the 3′ end of the message (data not shown) suggests that the the termination site is 42 bp downstream from the TAA stop codon. A striking stem-loop structure, reminiscent of procaryotic termination signals (Rosenberg and Court, 1979) as well as those found in other chloroplast genes (Dron *et al.*, 1982; Rochaix *et al.*, 1984), can be drawn from the sequence around this termination region. The stem contains 14 perfectly matched base pairs (Figure 3).

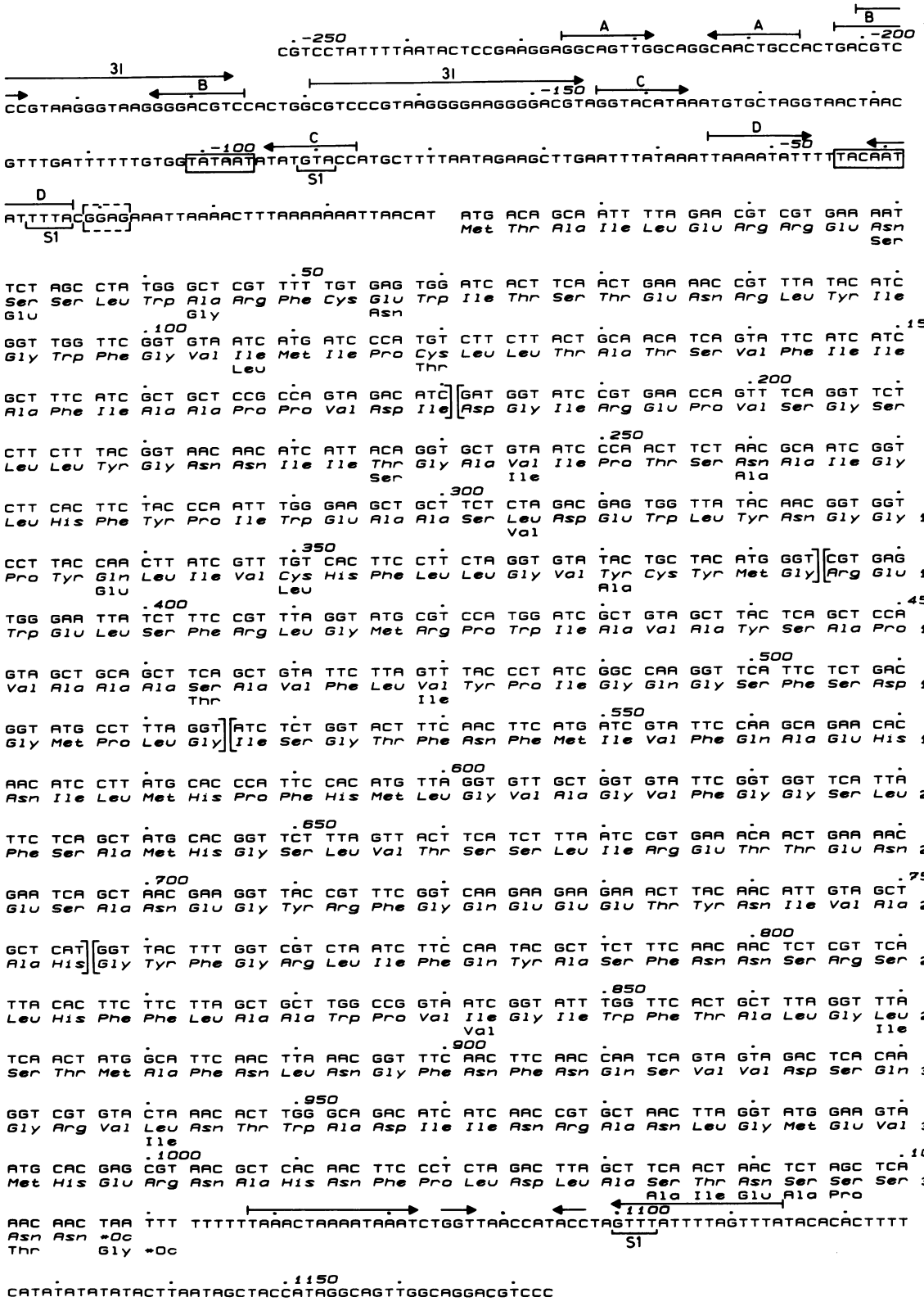

Fig. 3. DNA sequence of the coding regions of the *C. reinhardii psb*A gene and flanking nucleotides. The deduced amino acid sequence is given in italics below each codon, beginning with the first methionine in the open reading frame. Amino acid residues that differ in spinach (Zurawski *et al.*, 1982) are printed below the algal protein sequence. Brackets delineate the five exon regions of the *C. reinhardii* gene. Indicated in the 5′-non-coding sequence are a putative Shine-Dalgarno sequence (▨) (Shine and Dalgarno, 1974), two putative transcription initiation sites determined by S1 mapping (S1), and the Pribnow −10 sequence (▭) (Rosenberg and Court, 1979). Inverted repeats of eight or more bp are marked with arrows and labeled A−D. A direct repeat of 31 bp, with only two mismatches, is marked with arrows labeled 31. Indicated in the 3′-non-coding sequence is a large inverted repeat (→) and the S1 cleavage site marking the 3′ end of the mRNA. Nucleotides are numbered with +1 as the first base in the first methionine codon. Amino acid residues are numbered in the right margin.

Table I. Codon usage in *C. reinhardii psb*A

TTT-Phe 2	(12)[a]	TCT-Ser 12	(6)[a]	TAT-Tyr 0	(4)[a]	TGT-Cys 3	(5)[a]
TTC-Phe 24	(30)	TCC-Ser 0	(1)	TAC-Tyr 13	(6)	TGC-Cys 1	(0)
TTA-Leu 18	(21)	TCA-Ser 16	(8)	TAA-Ter 1	(1)	TGA-Ter 0	(0)
TTG-Leu 0	(1)	TCG-Ser 0	(1)	TAG-Ter 0	(0)	TGG-Trp 10	(13)
CTT-Leu 8	(7)	CCT-Pro 4	(5)	CAT-His 1	(1)	CGT-Arg 15	(12)
CTC-Leu 0	(0)	CCC-Pro 0	(0)	CAC-His 9	(8)	CGC-Arg 0	(1)
CTA-Leu 6	(4)	CCA-Pro 8	(7)	CAA-Gln 7	(10)	CGA-Arg 0	(0)
CTG-Leu 0	(0)	CCG-Pro 2	(0)	CAG-Gln 0	(2)	CGG-Arg 0	(0)
ATT-Ile 5	(16)	ACT-Thr 12	(9)	AAT-Asn 1	(1)	AGT-Ser 0	(4)
ATC-Ile 23	(4)	ACC-Thr 0	(0)	AAC-Asn 24	(14)	AGC-Ser 2	(0)
ATA-Ile 0	(2)	ACA-Thr 4	(10)	AAA-Lys 0	(3)	AGA-Arg 0	(0)
ATG-Met 12	(9)	ACG-Thr 0	(1)	AAG-Lys 0	(1)	AGG-Arg 0	(0)
GTT-Val 5	(13)	GCT-Ala 25	(27)	GAT-Asp 1	(3)	GGT-Gly 30	(29)
GTC-Val 0	(0)	GCC-Ala 0	(0)	GAC-Asp 6	(7)	GGC-Gly 1	(1)
GTA-Val 16	(9)	GCA-Ala 7	(9)	GAA-Glu 15	(9)	GGA-Gly 0	(0)
GTG-Val 0	(0)	GCG-Ala 0	(1)	GAG-Glu 4	(3)	GGG-Gly 0	(0)

[a]Codon usage in *C. reinhardii psb*D (Rochaix *et al.*, 1984).

Introns

Each exon-intron boundary indicated in Figure 3 was defined in two ways: by S1 mapping, and by comparison of the protein sequences deduced from the nucleotide sequences of the algal *psb*A and the non-interrupted spinach *psb*A. Fragments containing restriction sites located within the exon were ^{32}P end-labeled for S1 mapping of the 5′ and 3′ exon boundaries (sites indicated in Figure 2b). Fragments protected from S1 digestion after hybridization with *C. reinhardii* RNA were electrophoresed on sequencing gels alongside the nucleotide sequence of the initial labeled fragment, which allowed for the resolution of S1 fragment sizes within a few base pairs. In each case, the exon-intron junction predicted from the deduced protein sequences was consistent with the S1 mapping results. Because of the degeneracy of the genetic code, it is possible to shift the splicing positions in the first three introns by one base without changing the amino acid sequence. The precision of S1 mapping to within 1 – 3 bases does not permit a resolution of this duality. However, Figure 5 gives the most likely boundaries for the four introns, in which a T is found at the 3′ end of each exon preceding an intron, and a G is the last base of each intron.

This T. . .G configuration follows a splicing rule established for Group I introns (Michel *et al.*, 1982; Davies *et al.*, 1982). In addition, two different consensus sequences found in Group I introns are also found in *psb*A introns. The *box* 2 critical sequence (see Figure 6 legend), T.A.GA.ATAGTC, is found in introns II and IV, 32 bp and 67 bp upstream, respectively, from the following exon. Intron III contains two *box* 2 critical sequences at 25 and 100 bp 5′ to exon 4. The *box* 9 critical sequence, PyTCA. . .GACTA, is found in introns II and IV, 360 and 71 bp upstream, respectively, from the *box* 2 critical sequence. Intron I has been sequenced only 100 bp 5′ to exon 2. Sequence analysis of the *psb*A introns is not yet completed, but introns II and III contain open reading frames (ORF) coding for at least 44 and 43 amino acid residues, respectively. These ORFs are in-frame with and contiguous with the preceding exon reading frame. Figure 6 gives the overall organization of the *C. reinhardii psb*A gene, indicating the five exon regions coding for 60, 68, 47, 77 and 100 amino acids, respectively, the ORFs in introns II and III, and the location of the *box* 2 and *box* 9 critical sequences.

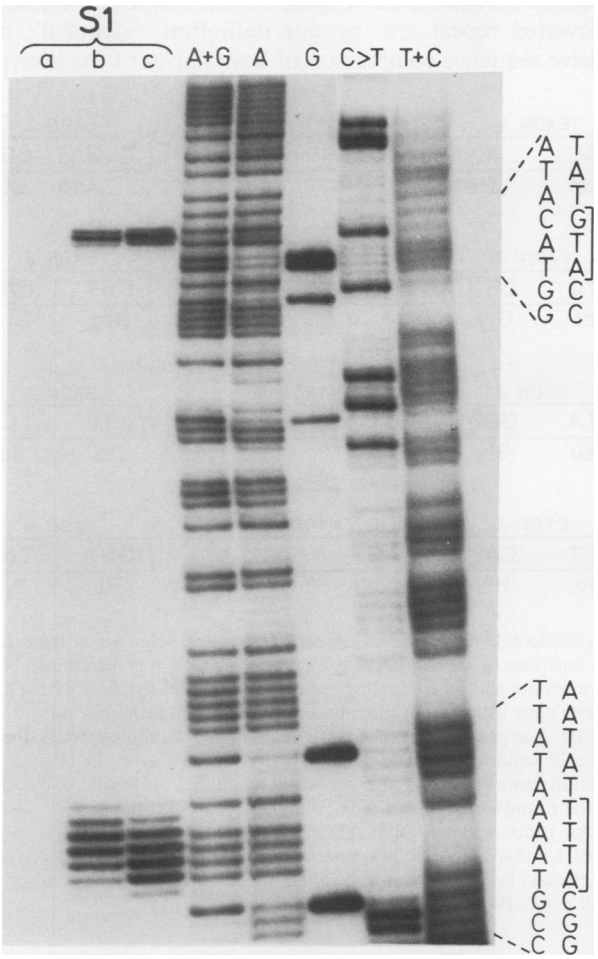

Fig. 4. Autoradiogram of an 8% urea-acrylamide sequencing gel showing S1 protection mapping at the 5′ end of *psb*A. A 500-bp *Sau*3A-*Dde*I fragment (Figure 2b) 5′-labeled at the *Sau*3A site (exon 1) was strand-separated for DNA chemical sequencing and S1 mapping. 20 μg of *C. reinhardii* RNA was hybridized with the labeled DNA (Materials and methods) and then digested with either 1500 U (S1 **lane b**) or 2500 U (**lane c**) of S1 nuclease. In the control (**lane a**), 20 μg of *E. coli* tRNA replaced the algal RNA. The Maxam-Gilbert sequence reactions performed are listed above each sequencing lane. S1 sites are indicated with brackets next to the relevant DNA sequences, assuming that an S1 protected fragment migrates 1.5 bp slower than its chemically cleaved counterpart (Sollner-Webb and Reeder, 1979).

Discussion

Analysis of the *C. reinhardii psb*A gene structure revealed two unexpected results. First, there are two *psb*A copies per circular chloroplast genome, as the gene is located entirely within the inverted repeat (Figure 1). Second, the *C. reinhardii psb*A is interrupted by four introns (Figure 2). These results are surprising in that the *psb*A genes from all higher plants examined to date are single copy genes, and none of them contain introns. The only genes previously identified in the chloroplast inverted repeat were rRNA and tRNA genes, but recent reports indicate that the gene *rp*19 for the ribosomal protein S19 has been found spanning one border of the soybean inverted repeat (Stutz, personal communication). In spinach, *rp*19 also spans one inverted repeat border, and the gene for the ribosomal protein L2 is found within the inverted repeat (Zurawski *et al.*, 1984).

psbA in the inverted repeat: ambiguities arising

The fact that the *C. reinhardii psb*A is found entirely within the inverted repeat raises questions about whether one can distinguish between the two gene copies. Sequences found in the inverted repeat are, by one definition, 'identical', but extensive sequence comparison of both halves of the inverted repeat has not been undertaken. Published results of sequences at the inverted repeat boundaries in tobacco report no differences between the two halves of the inverted repeat (Sugita *et al.*, 1984). We have taken care to clone *psb*A using restriction enzymes that do not cut within the gene, and span one boundary of the inverted repeat (Figure 1B). Thus, a given copy of *psb*A is identified by the presence of restriction site markers found in the adjacent single copy region. However, Palmer (1983) has proposed that recombination between sequences in the chloroplast inverted repeat is a common occurrence, resulting in an apparent flipping of one single copy region with respect to the other. Such flipping has recently been documented in *C. reinhardii* (Palmer, 1984). It is not yet known whether this recombination takes place throughout the entire inverted repeat or at one or more 'hot spots'. If recombination occurs within *psb*A itself or between *psb*A and the outside markers that define a given gene copy, then the definition we have used is simply operational. Even if we had cloned restriction fragments spanning the entire inverted repeat, the ambiguities in *psb*A would not be resolved (see Figure 7).

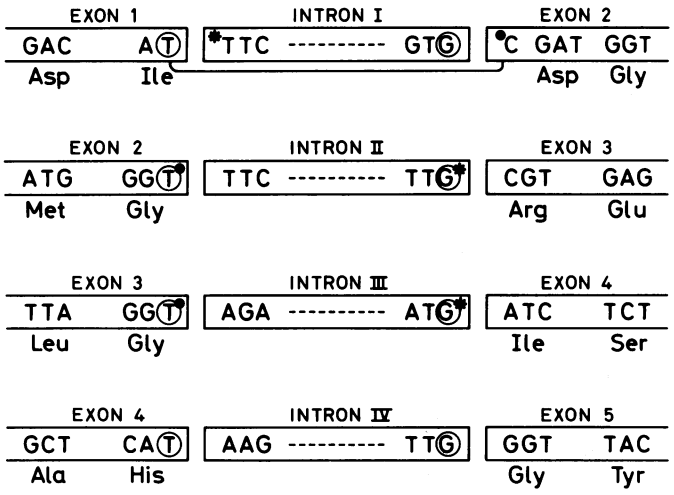

Fig. 5. Nucleotide sequence and deduced amino acid sequence at the exon-intron junctions of *C. reinhardii psb*A. Each junction was determined both by S1 mapping and by comparing the deduced protein sequence of the *C. reinhardii* gene with that of the uninterrupted spinach gene. For introns I, II and III, it is possible to splice at alternate points without affecting the amino acid sequence. In each case, the starred nucleotide in the intron would move into the adjacent exon, while the dotted nucleotide in the exon would be found in the intron. For example, this would change the isoleucine codon (exon 1) from ATC to ATT. Circled nucleotides demonstrate that a T is the last base in each exon preceding an intron, and a G is the last base of each intron, a feature common to Group I introns (Michel *et al.*, 1982).

psbA gene expression

In the light of the above ambiguities in defining a *psb*A copy, the question of *psb*A gene expression becomes complex. In the cyanobacterium *Anabaena* which contains at least three copies of *psb*A, S1 mapping and Northern analysis suggest that only one of these copies is expressed (Curtis and Haselkorn, 1984). However, in *C. reinhardii* the two copies of *psb*A, defined as discussed above, are indistinguishable by restriction analysis throughout the entire 7-kb gene, plus 1 kb of 5′-flanking sequences, and by DNA sequence analysis for the coding regions plus 400 bp of 5′- and 110 bp of 3′-flanking sequences. Moreover, sequence analysis of the junction between the inverted repeat and the single copy region near *psb*A (unpublished data) has shown that the end of the inverted repeat is ~1 kb 5′ to the *psb*A open reading frame. The size of the *psb*A message seen by Northern blot analysis is ~1.2 kb (Rochaix *et al.*, 1984). Thus, unless controlling regions >1 kb upstream of the structural gene affect its expression, or unless there are important single base pair

It is noted that in two different herbicide-resistant strains of *C. reinhardii* with different mutations in *psb*A (Erickson *et al.*, 1984 and unpublished data), both copies of *psb*A contain the given mutation. Unfortunately it is not clear at present whether this reflects the selective pressure for herbicide resistance, or the general recombination or gene conversion frequency (Palmer, 1983), or both. More needs to be known about the frequency of point mutations at a given locus, the 'dominant' or 'recessive' nature of the herbicide-resistant phenotype in these mutants, and the expression of both copies.

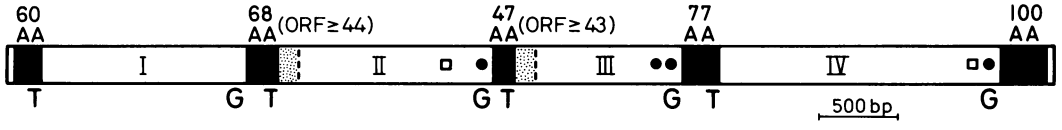

Fig. 6. Overall structure of *C. reinhardii psb*A. The five exons are marked with dark boxes and the number of amino acid residues (AA) per exon is noted. Size bar represents 500 bp. Stippled areas in introns II and III represent open reading frames of at least 44 and 43 amino acids, respectively, that are in frame with and contiguous with the preceding exon sequences. Characteristics these introns have in common with Group I introns (Michel *et al.*, 1982) are the G/T splicing rule (see Figure 5) and the presence of the *box* 2 (PyTCA.GACTA, ●) and *box* 9 (T.A.GA.ATAGTC, □) critical sequences. These sequences, identified by comparative sequence analysis of *cis*-dominant mutations in the yeast mitochondrial *box* locus, appear to be essential for proper intron splicing (De La Salle *et al.*, 1982). Group I introns have been identified in rRNA genes of fungi and protists, in one tRNA gene from higher plants (Bonnard *et al.*, 1984) and in various mitochondrial protein-encoding genes (Michel and Dujon, 1983).

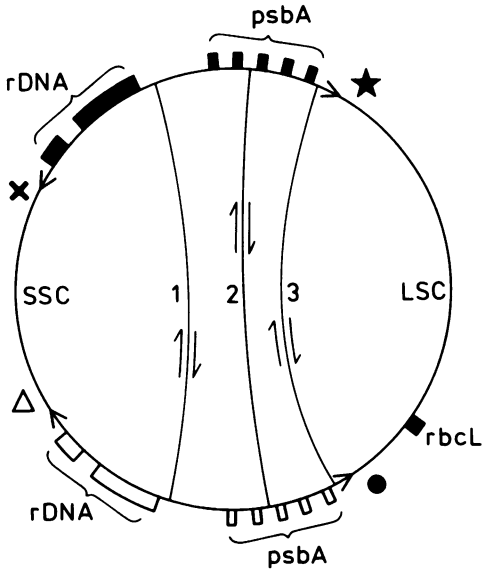

Fig. 7. Schematic map of the *C. reinhardii* chloroplast genome, showing possibilities for reciprocal recombination within the inverted repeat. The exons of *psb*A and the rDNA genes are represented by light boxes on one section of the inverted repeat and dark boxes on the other. Restriction site markers are indicated by symbols (★, ●, △, +) in the flanking large single copy region (LSC) and small single copy region (SSC). Reciprocal recombination at position 1 would not alter either copy of *psb*A with respect to the outside markers ★ and ●, while recombination at positions 2 and 3 would. The presence of pairs of outside markers (+ ★ or △ ●) on a given restriction fragment would rule out the occurrence of single but not double recombinational events. The position of *rbc*L, the gene coding for the large subunit of ribulose bisphosphate carboxylase, is given to orient the map.

changes or very small deletions in the 5′-flanking region between −400 and −1000 bp, it appears that both copies of the algal *psb*A gene are expressed. The only other deterrent to expression might be changes in the introns which have not been detected by restriction mapping or partial sequence analysis.

5′-flanking sequences

One feature of the *C. reinhardii psb*A 5′-flanking sequences, as also seen in the algal *rbc*L (Dron *et al.*, 1982) and *psb*D (Rochaix *et al.*, 1984) genes, is the presence of numerous inverted repeats and direct repeats. Figure 8 is a graphic depiction of the 5′ sequences of *psb*A, showing the location of both S1 cleavage sites in stem structures, and the corresponding Pribnow −10 sequences in the loops. The putative ribosomal binding site for the first AUG of the open reading frame is found just after the last stem loop structure. A very large, direct repeat of 31 bp with only two mismatches, is found covering one stem-loop structure, and spanning the region between it and the next stem-loop. It is not known whether the S1 cleavage sites reflect transcription initiation or mRNA processing. Possible folding of RNA into the two 5′ most stem-loop structures (Figure 8) gives a calculated free energy of −15 and −16 kcal/mol, respectively (Tinoco *et al.*, 1973). Similar calculations for the following two stem-loop structures give values of only −8.3 and −5 kcal/mol, respectively. Interestingly, a stem-loop structure and adjacent ribosome binding site is found in the 5′ *psb*A sequence of *Euglena* at approximately the same distance (−22 to −28 bp) from the first methionine codon (Karabin *et al.*, 1984).

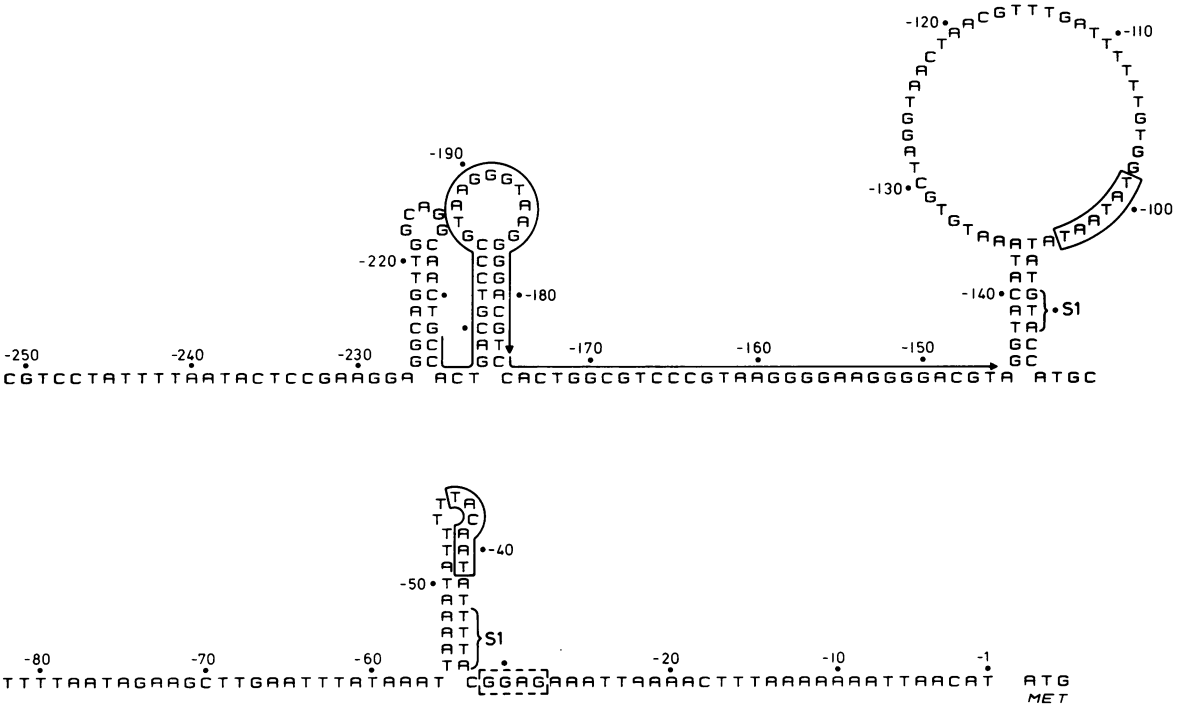

Fig. 8. Hypothetical secondary structure at the 5′ end of *psb*A, showing four stem-loop structures with at least 8 bp in the stem, two potential transcription initiation sites as determined by S1 mapping (S1), the corresponding Pribnow −10 sequences (▭), and a putative Shine-Dalgarno ribosome binding site (⁙). A direct repeat of 31 bp with only two mismatches is indicated with arrows. A 6-bp stem and loop structure could be drawn between residues −147 and −169. Bases are numbered negatively with respect to the first methionine in the open reading frame.

Introns

The presence of introns in the *C. reinhardii psb*A is surprising. None of the other protein genes examined in this alga appear to have introns, and *psb*A is uninterrupted in all higher plants examined to date. In contrast, electron microscopic examination of R-loop formation suggests that many protein genes in the chloroplast of the green alga *Euglena* contain introns (Koller and Delius, 1984). Sequence analysis of *Euglena tuf*A shows that it has introns (Montandon and Stutz, 1983) and the *Euglena* gene for the large subunit of ribulose bisphosphate carboxylase, *rbc*L, has nine introns (Koller *et al.*, 1984). The *Euglena psb*A has recently been sequenced, and, like that of *Chlamydomonas*, contains four introns (Keller and Stutz, 1984; Karabin *et al.*, 1984). However, these introns are smaller than those found in *C. reinhardii* and do not fall into the category of Group I introns as defined by Michel *et al.* (1982) and Davies *et al.* (1982). In addition, the introns interrupt the *Euglena* and *Chlamydomonas* coding sequence at different positions. The evolutionary significance of this is unclear, but is consistent with a different endosymbiotic origin of the *Euglena* chloroplast with respect to that of other plants and algae (Palmer, 1984).

The possible function of these introns in the algal chloroplast genes is not well understood. In *C. reinhardii*, the introns all fall into the category of Group I introns previously mentioned, having a well-defined secondary structure, G/T splicing junctions (Figure 5), and the conserved *box* 2 and *box* 9 critical sequences (Figure 6). In the light of the accumulating evidence for RNA 'maturases' (Jacq *et al.*, 1984), the presence of ORFs in two of the *C. reinhardii* introns (Figure 6) is intriguing. This situation is reminiscent of the yeast mitochondrial cytochrome b gene, in which exon-intron encoded RNA 'maturases' appear to be involved in some of the splicing reactions (De La Salle *et al.*, 1982). The *C. reinhardii* ORFs, coding for at least 43 and 44 amino acid residues, are in frame with the preceding exon sequences. Intron sequence data is incomplete, but in all other intron sequences compiled to date, the three reading frames are blocked, on average, every 12 amino acid residues. Interestingly, the 23S rRNA gene of *C. reinhardii* also contains an intron (Rochaix and Malnoe, 1978) which falls into the Group I classification, and this intron has an ORF (unpublished data). So far, ORFs encoding 'maturases' have been identified in the mitochondrial genome of several species of fungi, but not yet any chloroplast or nuclear genes. Further work is necessary to determine the functional importance, if any, of these *C. reinhardii* chloroplast introns, and their possible role as maturases.

Structure and function of the psbA gene product

The exact size of the *psb*A gene product remains a mystery: there is difficulty in accurately estimating the mol. wt. of hydrophobic, membrane proteins which are known to migrate anomalously in various gel systems (Darley-Usmar and Fuller, 1981; Williams *et al.*, 1983; Youvan *et al.*, 1984), and DNA sequence analysis shows the presence of two ATGs which may serve as initiation codons. We have indicated (Figure 3) the sequence for the largest, 38.95-kd protein coded for by *psb*A, as was given for spinach and *N. debneyi* (Zurawski *et al.*, 1982). While it is possible that this protein starts at the second methionine (Hirschberg *et al.*, 1984), we see a conservation of both nucleotide and amino acid sequence beginning with the first methionine, suggesting a selective pressure on these sequences. DNA sequence homology between the alga and higher plants is ~75% for the 108 bp between the first two methionine codons, and this level continues throughout the coding region of exons 1 and 2. In exon 3, the homology is 78%, while both exon 4 and exon 5 up to codon 344 are 84% homologous with the spinach gene. This DNA homology is all the more striking since the 5′ and 3′ non-translated sequences in the *C. reinhardii* gene have very little homology with the corresponding region in the higher plant genes. A Shine-Dalgarno like sequence (GGTGT) can be found 4 bp 5′ to the second methionine in both *C. reinhardii* and higher plants. Another possible ribosome binding site (GGAG) can be found 5′ to the first methionine by 27 bp in the alga and 35 bp in spinach.

Analysis of higher plant *psb*A genes examined to date (spinach, tobacco, pigweed, soybean, maize, mustard) shows that the protein is highly conserved. Sequence analysis of *psb*A from *C. reinhardii* indicates that this conservation extends even to algae. Much of the divergence occurs at the carboxyl terminus, which appears to be removed during processing (Marder *et al.*, 1984). For amino acid residues 1 – 344, the *C. reinhardii* protein is 95% homologous to that of higher plants, and 88% and 87% homologous to the corresponding *Euglena* and *Anabaena* proteins, respectively. The amino terminus of the protein is also divergent, with an 86% homology of *C. reinhardii* to spinach between the first two methionine residues. More than one third of the protein, from residues 161 to 280, is 100% conserved between *C. reinhardii* and the higher plants. Within this region is a stretch of 26 amino acid residues, (residues 196 – 221) identified by Hearst and Sauer (1984), which show significant homology with reaction center subunits L and M from the purple photosynthetic bacteria (Williams *et al.*, 1983; Youan *et al.*, 1984). We also see a similar region of homology in the deduced protein sequence of D2, another PSII thylakoid membrane protein from *C. reinhardii* (Rochaix *et al.*, 1984). Thus, these amino acid residues may constitute an important functional region of the protein, with respect to electron transport, and may in fact determine quinone binding. An azido derivative of the s-triazine, atrazine, which appears to compete with quinone binding (Vermaas *et al.*, 1983), binds to a tryptic digest fragment of the 32-kd protein which includes this highly conserved region (Wolber and Steinback, 1984). Models based on hydrophobicity predict that the 32-kd protein contains several membrane spanning helices (Rao *et al.*, 1983), one of which includes the 26 amino acid conserved region. In addition, sequence analysis of *psb*A from herbicide-resistant strains of *C. reinhardii* and higher plants shows that the mutations in *psb*A alter amino acid residues that lie within (Erickson, Mets, Rochaix, unpublished data) or in the membrane-spanning region adjacent to (Erickson *et al.*, 1984; Hirschberg *et al.*, 1984) this 26 amino acid region. Further studies on quinone binding, herbicide binding and analysis of various different herbicide-resistant *C. reinhardii* mutants and their revertants, should reveal much about the function of this *psb*A gene product and its role in PSII electron transport.

Materials and methods

Algal and bacterial strains

C. reinhardii strains *cw*15 (a cell-wall mutant, wild-type phenotype with respect to herbicide resistance, Davies and Plaskitt, 1971) and DCMU4 (a herbicide-resistant mutant of wild-type strain 137c, Erickson *et al.*, 1984) were maintained in continuous dim light on solid tris-acetate-phosphate (TAP) medium (Gorman and Levine, 1965). *E. coli* C600 (r_k^-, m_k^-, SuII) was the

host for propagation of lambda phage and plasmids. *E. coli* NM539 [hsdR supF trpR met (P2 cox3)] was obtained from N.Murray and used to select for recombinant phage. *In vitro* packaging extracts were prepared from *E. coli* strains BHB2690 [N205 recA-(λimm434 CI^{ts} b2 red3 Dam15 Sam7)/λ] and BHB2688 [N205 recA-(λimm434 CI^{ts} b2 red3 Eam15 Sam7)/λ].

DNA preparation

Chloroplast DNA was prepared from *C. reinhardii* strains *cw*15 as previously described (Rochaix, 1980). DNA from DCMU4 was prepared in essentially the same manner, except that pelleted cells were frozen in a dry-ice ethanol bath and quick-thawed twice before re-suspension and pronase digestion. Large scale preparations of plasmid DNA and lambda DNA were prepared as described previously (Erickson *et al.*, 1984). Rapid preparations of plasmid DNA were performed as described by Holmes and Quigley (1981), except that 1.5 ml of cells were pelleted in an Eppendorf tube and re-suspended in 105 μl of the prescribed sucrose solution. 7.5 μl of lysozyme (10 mg/ml) was added, samples were set at room temperature for 5 min, and then boiled for 1 min.

Cloning

The lambda vector λEMBL4 (Frischauf *et al.*, 1983) was used to clone *Bam*HI or *Bam*HI/*Bgl*II fragments from *C. reinhardii* chloroplast DNA. Ligated DNA was packaged *in vitro* (Maniatis *et al.*, 1982) and the resulting phage were infected into *E. coli* strain NM539, which prevents growth of parental phage. Recombinant plaques were screened by the method of Benton and Davis (1977), using the spinach *psbA* gene (1.2-kb *Xba*I/*Sma*I fragment from pSOCS15, Zurawski *et al.*, 1982) as a probe. Hybridizing phage were plaque purified several times on *E. coli* NM539 before large scale propagation in *E. coli* C600.

To facilitate sequence analysis, chloroplast DNA contained in the lambda recombinants was subcloned into plasmid vectors as follows. The 2.5-kb *Hind*III/*Bam*HI fragment containing exon 5 (see Figure 2b) was subcloned into pBR322 (Bolivar *et al.*, 1977). *Eco*RI restriction fragments R14 and R16, containing *psbA* exons 1 – 4, were subcloned into pBR328 (Covarrubias *et al.*, 1981). The 5′ end of the gene contained in R16 was subcloned in pBR322 as a 1.4-kb *Eco*RI/*Hind*III fragment (see Figure 2b). The chloroplast *Bam*HI fragment Ba12, from *cw*15, had been previously cloned in the plasmid vector pBR313 (Rochaix, 1978). R14 and the 2.5-kb *Hind*III/*Bam*HI fragment from this *cw*15 Ba12 were likewise subcloned into the plasmid vectors as described above. Recombinant plasmids were identified by their sensitivity to tetracycline (pBR322) or chloramphenicol (pBR328). Conditions for ligations, transformations, nick-translation, hybridization and autoradiography are as previously described (Erickson *et al.*, 1981).

DNA sequence analysis

DNA fragments, electro-eluted from acrylamide gels and digested with the desired restriction enzymes, were labeled either at the 3′ end with an appropriate [α-^{32}P]dNTP and the large fragment of DNA polymerase I (Klenow fragment), or at the 5′ end (after treatment with calf-intestine alkaline phosphatase) with [γ-^{32}P]ATP and T4 polynucleotide kinase. Fragments were strand-separated by electrophoresis on an acrylamide gel, or were re-digested with another enzyme before electrophoresis on a non-denaturing acrylamide gel. Recovered fragments, labeled at one end only, were sequenced by the chemical cleavage technique of Maxam and Gilbert (1980). Reaction products were separated by electrophoresis on 8% and 20% urea-acrylamide gels, and visualized by autoradiography. A Hewlett Packard computer, model 9845, was used to analyse DNA sequences.

S1 mapping

S1 nuclease mapping was performed according to Berk and Sharp (1977). DNA fragments were 5′ and 3′ end-labeled as described above, and strand separated before hybridization with total *C. reinhardii* RNA. Labeled fragments used for mapping of the intron/exon borders and the 5′ and 3′ end of the transcripts are indicated in Figure 2b.

Enzymes, isotopes

Restriction endonucleases, the large fragment of DNA polymerase I (Klenow), and T4 polynucleotide kinase were purchased from Genofit (Geneva), Bethesda Research Laboratories, Inc., Boehringer Mannheim or Amersham and used as recommended by the supplier. Calf-intestine alkaline phosphatase was purchased from Boehringer Mannheim, and radioisotopes were purchased from Amersham.

Acknowledgements

We thank Hans Bohnert for sending us the spinach *psbA* to use as a probe, N.Murray for strains NM539 and λEMBL4, O.Jenni for drawings and photography, and M.Goldschmidt-Clermont for critical reading of the manuscript. This work was supported by the Swiss National Science Foundation, grant No 3.258.0.82.

References

Arntzen,C.J., Pfister,K. and Steinback,K.E. (1982) in LeBaron,H.M. and Gressel,J. (eds.), *Herbicide Resistance in Plants*, Wiley-Interscience Publications, NY, pp. 185-214.
Bedbrook,J.R., Link,G., Coen,D., Bogorad,L. and Rich,A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3060-3064.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Bolivar,F., Rodriguez,R., Greene,P., Betlach,M., Heyneker,H. and Boyer, H.W. (1977) *Gene*, **2**, 95-113.
Bonnard,G., Michel,F., Weil,J. and Steinmetz,A. (1984) *Mol. Gen. Genet.*, **194**, 330-336.
Bottomley,W., Spencer,D. and Whitfield,P.R. (1974) *Arch. Biochem. Biophys.*, **164**, 106-117.
Chua,N.H. and Gillham,N.W. (1977) *J. Cell Biol.*, **74**, 441-452.
Covarrubias,L., Cervantes,L., Covarrubias,A., Soberon,X., Vichido,I., Blanco,A., Kupersztoch-Portnoy,K. and Bolivar,F. (1981) *Gene*, **13**, 25-35.
Curtis,S. and Haselkorn,R. (1984) *Plant Mol. Biol.*, **3**, 249-258.
Darley-Usmar,M. and Fuller,S.D. (1981) *FEBS Lett.*, **135**, 164-166.
Davies,D.R. and Plaskitt,A. (1971) *Methods Enzymol.*, **21**, 413-428.
Davies,R.W., Waring,R., Ray,J., Brown,T. and Scazzocchio,C. (1982) *Nature*, **300**, 719-724.
De La Salle,H., Jacq,C. and Slonimski,P. (1982) *Cell*, **28**, 721-732.
Delepelaire,P. (1983) *Photochem. Photobiophys.*, **6**, 279-292.
Delepelaire,P. (1984) *EMBO J.*, **3**, 701-706.
Dron,M., Rahire,M. and Rochaix,J.D. (1982) *J. Mol. Biol.*, **162**, 775-793.
Eaglesham,A. and Ellis,R.J. (1974) *Biochim. Biophys. Act*, **335**, 396-407.
Edelman,M. and Reisfeld,A. (1980) in Leaver,C. (ed.), *Genome Organization and Expression in Plants*, Plenum Press, NY, pp. 353-362.
Erickson,J.M., Rushford,C.L., Dorney,D.J., Wilson,G.N. and Schmickel, R.D. (1981) *Gene*, **16**, 1-9.
Erickson,J.M., Rahire,M., Bennoun,P., Delepelaire,P., Diner,B. and Rochaix,J.D. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3617-3621.
Frischauf,A., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Gorman,D.S. and Levine,R.P. (1965) *Proc. Natl. Acad. Sci. USA*, **54**, 1665-1669.
Grebanier,A., Coen,D., Rich,A. and Bogorad,L. (1978) *J. Cell Biol.*, **78**, 734-746.
Hearst,J.E. and Sauer,K. (1984) *Z. Naturforsch.*, **39C**, 421-424.
Hirschberg,J. and McIntosh,L. (1983) *Science (Wash.)*, **222**, 1346-1348.
Hirschberg,J., Bleeker,A., Kyle,D., McIntosh,L. and Arntzen,C.J. (1974) *Z. Naturforsch.*, **39C**, 412-420.
Hoffman-Falk,H., Mattoo,A.K., Marder,J.B., Edelman,M. and Ellis,R.J. (1982) *J. Biol. Chem.*, **257**, 4583-4587.
Holmes,D. and Quigley,M. (1981) *Anal. Biochem.*, **114**, 193.
Jacq,C., Banroques,J., Becam,A., Slonimski,P., Guiso,N. and Danchin,A. (1984) *EMBO J.*, **3**, 1567-1572.
Karabin,G., Farley,M. and Hallick,R.B. (1984) *Nucleic Acids Res.*, **12**, 5801-5812.
Keller,M. and Stutz,E. (1984) *FEBS Lett.*, **175**, 173-177.
Koller,B., Gingrich,J., Farley,M., Delius,H. and Hallick,R.B. (1984) *Cell*, **36**, 545-553.
Koller,B. and Delius,H. (1984) *Cell*, **36**, 613-622.
Kuwabara,T. and Murata,N. (1979) *Biochim. Biophys. Acta*, **581**, 228-236.
Link,G. and Langridge,U. (1984) *Nucleic Acids Res.*, **10**, 593-605.
Malnoe,P., Rochaix,J.D., Chua,N.H. and Spahr,P.F. (1979) *J. Mol. Biol.*, **133**, 417-434.
Maniatis,T., Fritsch,E. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Marder,J.B., Goloubinoff,P. and Edelman,M. (1984) *J. Biol. Chem.*, **259**, 3900-3908.
Mattoo,A., Pick,U., Hoffman-Falk,H. and Edelman,M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1572-1576.
Mattoo,A., Hoffman-Falk,H., Marder,J.B. and Edelman,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1380-1384.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Michel,F., Jacquier,A. and Dujon,B. (1982) *Biochimie*, **64**, 867-881.
Michel,F. and Dujon,B. (1983) *EMBO J.*, **2**, 33-38.
Montandon,P. and Stutz,E. (1983) *Nucleic Acids Res.*, **11**, 5877-5892.
Mulligan,B., Schultes,N., Chen,L. and Bogorad,L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2693-2697.
Ohad,I., Kyle,D. and Arntzen,C.J. (1984) *J. Cell Biol.*, **99**, 481-485.
Palmer,J.D. (1983) *Nature*, **301**, 92-93.
Palmer,J.D. (1984) in MacIntyre,R.J. (ed.), *Monographs in Evolutionary*

Biology: Molecular Evolutionary Genetics, Plenum Publishing, NY, in press.
Pfister,K., Steinback,K., Gardner,G. and Arntzen,C.J. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 981-985.
Rao,J.K.M., Hargrave,P.A. and Argos,P. (1983) *FEBS Lett.,* **156**, 165-169.
Reisfeld,A., Mattoo,A. and Edelman,M. (1982) *Eur. J. Biochem.,* **124**, 125-129.
Rochaix,J.D. (1978) *J. Mol. Biol.,* **126**, 597-617.
Rochaix,J.D. (1980) *Methods Enzymol.,* **65**, 785-795.
Rochaix,J.D. and Malnoe,P. (1978) *Cell,* **15**, 661-670.
Rochaix,J.D., Dron,M., Rahire,M. and Malnoe,P. (1984) *Plant Mol. Biol.,* **3**, 363-370.
Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.,* **13**, 319-353.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA,* **71**, 1342-1346.
Smith,H.O. and Birnstiel,M. (1976) *Nucleic Acids Res.,* **3**, 2387-2398.
Spielmann,A. and Stutz,E. (1983) *Nucleic Acids Res.,* **11**, 7157-7167.
Sollner-Webb,B. and Reeder,R.H. (1979) *Cell,* **18**, 485-499.
Steinback,K., McIntosh,L., Arntzen,C.J. and Bogorad,L. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 7463-7467.
Sugita,M., Kato,A., Shimada,H. and Sugiura,M. (1984) *Mol. Gen. Genet.,* **194**, 200-205.
Sugita,M. and Sugiura,M. (1984) *Mol. Gen. Genet.,* **195**, 308-313.
Tinoco,I., Borer,P.N., Dengler,B., Levine,M., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature New Biol.,* **246**, 40-41.
Trebst,A. (1980) *Methods Enzymol.,* **69**, 675-715.
Vermaas,W., Arntzen,C., Gu,L.-Q. and Yu,C.-A. (1983) *Biochim. Biophys. Acta,* **723**, 266-275.
Vermaas,W., Steinback,K. and Arntzen,C.J. (1984) *Arch. Biochem. Biophys.,* **231**, 226-232.
Velthuys,B.R., Amesz,J. (1974) *Biochim. Biophys. Acta,* **333**, 85-94.
Wettern,M., Owens,J.C. and Ohad,I. (1983) *Methods Enzymol.,* **97**, 554-567.
Williams,J., Steiner,L., Ogden,R., Simon,M. and Feher,G. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 6505-6509.
Wolber,P.K. and Steinback,K.E. (1984) *Z. Naturforsch.,* **39C**, 424-429.
Youvan,D., Bylina,E., Alberti,M., Begusch,H. and Hearst,J.E. (1984) *Cell,* **37**, 949-957.
Zurawski,G., Bohnert,H., Whitfeld,P. and Bottomley,W. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 7699-7703.
Zurawski,G. Bottomley,W. and Whitfield,P. (1984) *Nucleic Acids Res.,* **12**, 6547-6558.

Received on 22 August 1984