Mol Gen Genet (1986) 202:1–5

MGG

Intron in the gene for the ribosomal protein S16 of tobacco chloroplast and its conserved boundary sequences

Kazuo Shinozaki[1], Hiroshi Deno[2], Mamoru Sugita[3], Seiki Kuramitsu[4], and Masahiro Sugiura[1]

[1] Center for Gene Research, Nagoya University, Chikusa, Nagoya 464
[2] Bioscience Research Center, Mitsui Petrochemical Industries, Waki-cho, Kuga-gun, Yamaguchi-ken 740
[3] Department of Botany, Hokkaido University, Sapporo 060
[4] Department of Medical Chemistry, Osaka Medical College, Takatsuki, Osaka 569, Japan

Summary. The nucleotide sequence of a tobacco chloroplast gene coding for a protein homologous to *Escherichia coli* ribosomal protein S16 (*rps16*) has been determined. It is located 690 bp upstream from the gene for $tRNA^{Lys}$ (UUU) on the same DNA strand. This gene has an 860 bp intron. The intron boundary sequences are similar to those of the tobacco chloroplast split genes for $tRNA^{Gly}$ (UCC), $tRNA^{Lys}$ (UUU), $tRNA^{Val}$(UAC) and ribosomal proteins L2 and S12. We classified these introns as a new group, designated group III, which is different from groups I and II proposed by Michel and Dujon (1983). The conserved boundary sequences of the group III introns are GTGCGNY at the 5′ ends and ATCNRY(N)YYAY at the 3′ ends. The *rps16* coding region contains 255 bp (85 codons) and its deduced amino acid sequence shows 34% homology (46% conservative replacements) with that of *E. coli* ribosomal protein S16. The primary transcript of *rps16* is 1.3 kb long.

Key words: DNA sequence – Chloroplast – Ribosomal protein gene – Intron – Splicing mechanism

Introduction

Chloroplasts contain their own transcriptional and translational systems. Chloroplast ribosomes are 70S in size and contain 4 to 5 rRNAs and about 60 ribosomal protein species (see Whitfeld and Bottomley 1983). All the chloroplast rRNAs are known to be encoded in chloroplast DNA (see Whitfeld and Bottomley 1983) and about one-third of chloroplast ribosomal proteins are thought to be encoded in chloroplast DNA (Eneas-Filho et al. 1981). Several of the genes for chloroplast ribosomal proteins (designated CS or CL; Sugita and Sugiura 1983) have been cloned and sequenced: *rps19*, the gene coding for CS19, in *Nicotiana tabacum* (Sugita and Sugiura 1983); *rps4*, the gene for CS4, in maize (Subramanian et al. 1983); *rps19* and *rpl2*, the genes for CS19 and CL2, respectively, in spinach and *Nicotiana debneyi* (Zurawski et al. 1984) *rps12* and *rps7*, the genes for CS12 and CS7, respectively, in *Euglena* (Montandon and Stutz 1984) and *rps14*, the gene for CS14, in liverwort (Umesono et al. 1984). The *N. debneyi rpl2* gene contains a long intron (Zurawski et al. 1984). The deduced amino acid sequences of CS19, CS4, CL2, CS12, CS7 and CS14 as well as the determined amino acid sequence of spinach CL12 (Bartsch et al. 1982) show moderate homology (40%–60%) with those of corresponding *Escherichia coli* ribosomal proteins.

Here we present the nucleotide sequence of a cloned fragment of tobacco chloroplast DNA and show that it encodes a protein homologous to *E. coli* ribosomal protein S16. This gene (*rps16*) is interrupted by an 860 bp sequence. The intron boundary sequences of the chloroplast genes in higher plants analysed so far have been compiled and the conserved sequences are discussed in relation to molecular mechanisms of splicing in chloroplasts.

Offprint requests to: K. Shinozaki

Materials and methods

Recombinant plasmid pTB25, which contains a 15.7 kb *Bam*HI partial fragment of *Nicotiana tabacum* (var. Bright Yellow 4) chloroplast DNA, was constructed as described using pBR322 (Sugiura and Kusuda 1979). DNA sequencing was performed by the chemical method (Maxam and Gilbert 1977) and the dideoxy chain termination method (Sanger et al. 1977). M13 vectors mp10 and mp11 and *E. coli* JM109 were used for cloning (Messing et al. 1981). DNA sequences were analysed using the GENETYX software system developed by Software Development Co. Ltd., Tokyo, Japan. Northern blot hybridization was carried out as described (Ohme et al. 1985). Tobacco chloroplasts were prepared from young leaves grown in a greenhouse or stored in a dark room for 2 days after growth in a greenhouse. Tobacco chloroplast RNA was extracted as described (Shinozaki et al. 1983).

Results and discussion

The DNA sequence

We have cloned tobacco chloroplast DNA fragments produced by partial digestion with *Bam*HI followed by size fractionation (Sugiura and Kusuda 1979). One of the recombinant plasmids, pTB25, contains 2.9, 0.45, 3.5, 0.17 and 8.7 kb *Bam*HI fragments in this order in pBR322 (Fig. 1). The inserted sequence has been partialy sequenced and shown to contain the 5′ half of *trnK*, the gene for $tRNA^{Lys}$ (UUU) (Sugita et al. 1985), *trnQ*, *trnS*, *trnG* and *trnR*, the genes for $tRNA^{Gln}$ (UUG), $tRNA^{Ser}$ (GCU), $tRNA^{Gly}$ (UCC) and $tRNA^{Arg}$ (UCU), respectively (Deno and Sugiura 1983, 1984) and *atpA*, *atpF* and *atpH*, the genes for the α, I and III subunits of proton-translocating ATPase, respectively (Deno et al. 1983, 1984; Fig. 1). We se-

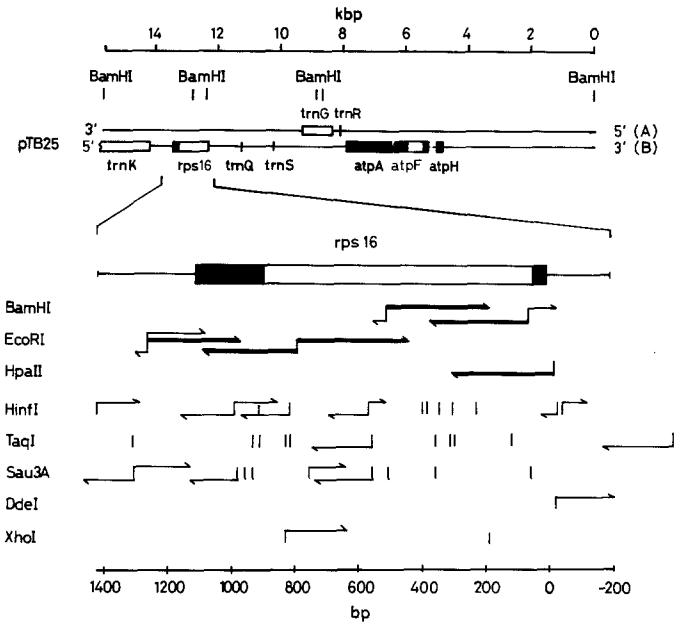

Fig. 1. Physical map of the cloned 15.7 kb partial *Bam*HI fragment from tobacco chloroplast DNA and the strategy for sequencing part of it. Strand A (*A*) contains *trnQ* and *trnR*. Strand B (*B*) contains *atpH, atpF, atpA, trnS, trnQ, rps16* and half of *trnK*. Coding regions are shown by *thick lines* and introns by *boxes*. The *lower part* shows an expanded physical map of *rps16*. *Vertical lines* indicate restriction sites. *Horizontal thin arrows* and *bold arrows* show the direciton and extent of the DNA segments analysed by the chemical method and the dideoxy chain termination method, respectively

quenced the remaining 1.6 kb portion between *trnK* and *trnQ* in the consecutive 2.9, 0.45 and 3.5 kb *Bam*HI fragments by the strategy shown in Fig. 1. Figure 2 shows the sequence of a 1,626 bp portion.

The gene for a protein homologous to E. coli S16

A short open reading frame (ORF_{14}), in which the deduced sequence of the first 14 amino acid residues is similar to the NH_2-terminus of *E. coli* ribosomal protein S16, was found at positions 1 to 42 and was 382 bp downstream from *trnQ* in the same orientation in the 3.5 kb *Bam*HI fragment (Fig. 2). We then searched for a reading frame showing sequence homology with the rest of *E. coli* S16 and found one of 71 codons (ORF_{71}, positions 903 to 1,115) 860 bp downstream of ORF_{14} and 690 bp upstream of *trnK* on the same strand. ORF_{14} plus ORF_{71} correspond to 85 amino acid residues while *E. coli* S16 is 82 residues long (Vandekerckhove et al. 1977). In Fig. 3 amino acid sequences deduced from $ORF_{14}+ORF_{71}$ and of *E. coli* S16 are aligned (with a few gaps to maximize homology) and homologous residues are boxed. Twenty-nine identical amino acid residues are found out of the 85 positions (i.e. 34%) and 10 conservative replacements (i.e. 12%), giving an overall homology of 46% between the two sequences. Homologous residues in *E. coli* S16 and $ORF_{14}+ORF_{71}$ are distributed throughout the polypeptides and their conserved secondary structures, predicted by the method of Chou and Fasman (1978), gave essentially the same profiles (Fig. 3). The amino acid sequences of CL12, CS19, CS4, CL2 and CS14 showed 50%, 55%, 39%, 40% and 45% homology, respectively, with the *E. coli* counterparts (Bartsch et al. 1982; Sugita and Sugiura 1983; Subramanian et al. 1983; Zurawski et al. 1984; Umesono et al. 1984). Based on the primary and secondary structural homologies, we concluded that $ORF_{14}+ORF_{71}$ represents the gene for tobacco chloroplast ribosomal protein CS16 (*rps16*). The *rps16* coding region contains 255 bp (85 codons) and the molecular weight of the tobacco CS16 deduced from the DNA sequence is 9,921 daltons.

The intron in rps16

Allocation of the intron site in tobacco *rps16* was determined by comparing the intron-exon boundary sequence of *rps16* with those of the tobacco *trnG, trnK, rpl2* and *rps12* genes (Hallick et al. 1985; Sugita et al. 1985). The boundary sequences of these reported introns have a conserved sequence (GTGCGNY) at the 5′ ends and a conserved sequence (ATCNRYY(N)YYAY) at the 3′ ends (Fig. 4). We found a GTGCGAC sequence (positions 41 to 47) in *rps16* which coincides with the 5′ conserved boundary sequence. Therefore, the *rps16* intron most likely starts at position 41 (Fig. 2). To follow the 3′ conserved sequence and produce the 14^{th} arginine codon (CGN), the intron should end at position 900. The intron is hence 860 bp long. No significant open reading frames were found in either strand of the 860 bp intron.

We have preliminarily proposed that introns found in chloroplast genes can be classified into three groups (Sugita et al. 1985). We compiled the intron boundary sequences of the chloroplast genes reported so far (Fig. 4) and also compared them with the nuclear and mitochondrial intron sequences to confirm our previous proposal.

Group I introns (Michel and Dujon 1983). The intron of *trnL*, the gene for $tRNA^{Leu}$ (UAA) (Steinmetz et al. 1982; Bonnard et al. 1984; unpublished observation), can be folded into a secondary structure which is similar to the postulated structure of the intron of the autospliceable rRNA precursor of *Tetrahymena* (Bonnard et al. 1984). A T at the 3′ ends of their 5′ exons and a G at the 3′ ends of their introns are common to all group I introns including *Tetrahymena* rDNA (Cech 1983). Group I introns of mitochondrial genes in lower eukaryotes have additional conserved sequences – named box2 and box9, sequences A and B (Michel and Dujon 1983). Open reading frames encoding "maturases" have been proposed for the mitochondrial genomes of lower eukaryotes. Sequences similar to box2 and box9 were found in chloroplast *trnL* introns, although no significant open reading frames were found. Recently introns in the gene for the 32 kd protein (Erickson et al. 1984) and 23 S rRNA in *Chlamydomonas* (Rochaix et al. 1985) were found to belong to group I and contain "maturase"-like open reading frames. The chloroplast group I introns may be spliced out by an autocatalytic reaction, as is the case for the *Tetrahymena* rRNA precursor, or by a "maturase" as for cytochrome b and cytochrome oxidase mRNA precursors.

Group II introns (Michel and Dujon 1983). Introns of *trnA* and *trnI*, the genes for $tRNA^{Ala}$ (UGC) and $tRNA^{Ile}$ (GAU), respectively (Koch et al. 1981; Takaiwa and Sugiura 1982), can be folded into a secondary structure which is similar to the postulated structure of introns of the genes for maize mitochondrial cytochrome oxidase and yeast cytochrome oxidase and cytochrome b (Michel and Dujon

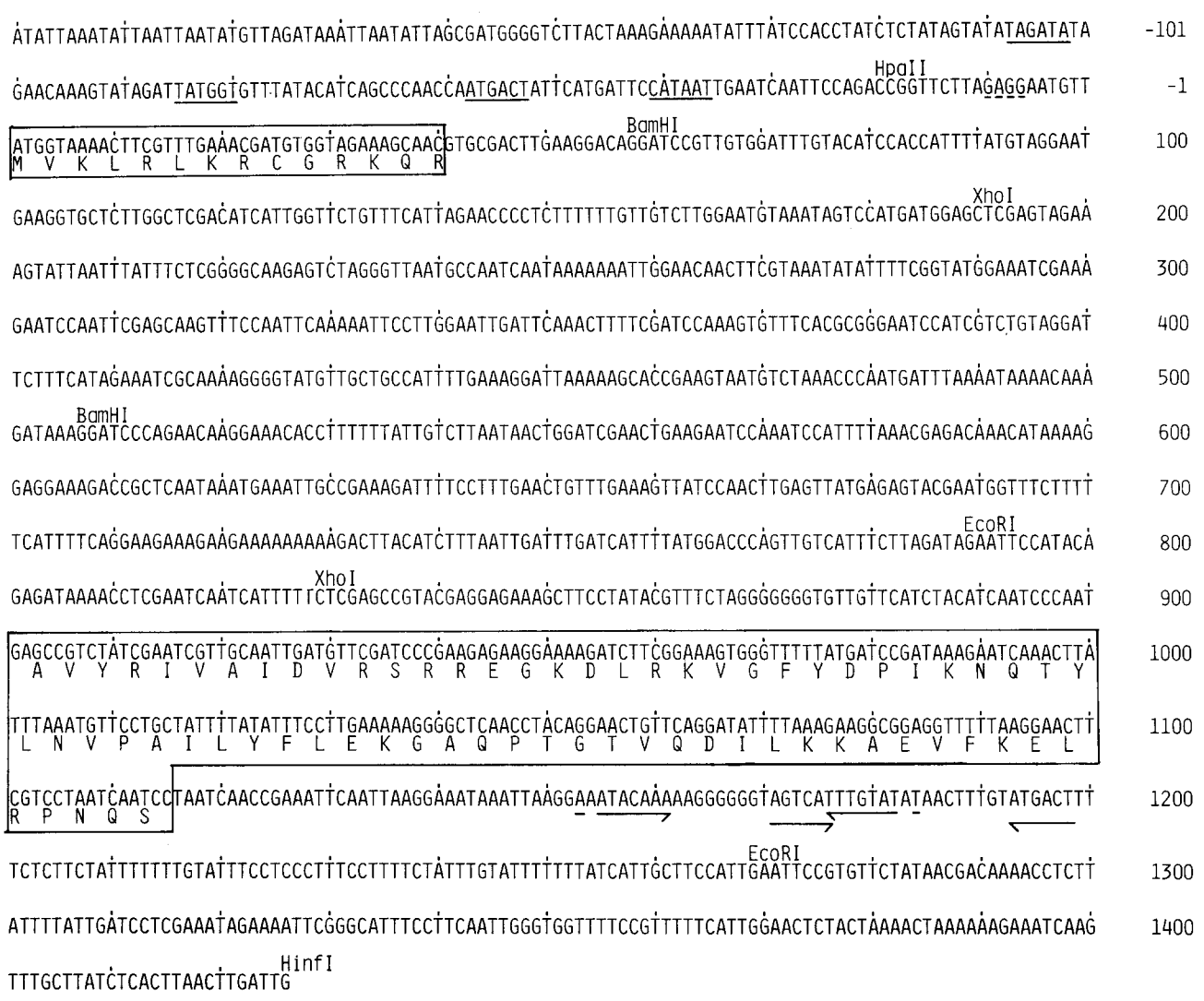

Fig. 2. DNA sequence of the 1,626 bp region containing *rps16*. The RNA-like strand (strand A) is presented. Exons are *boxed*. The deduced amino acid sequence is shown below the DNA sequence. Sequences similar to the "Pribnow box" (−10) and "−35 region" (−35) are *underlined*. A *broken line* indicates the Shine-Dalgarno (SD) sequence. *Horizontal arrows* indicate inverted repeats. The *Hin*fI site at position 1,426 of this sequence is the same as the *Hin*fI site at position −378 of the sequence reported previously (Sugita et al. 1985)

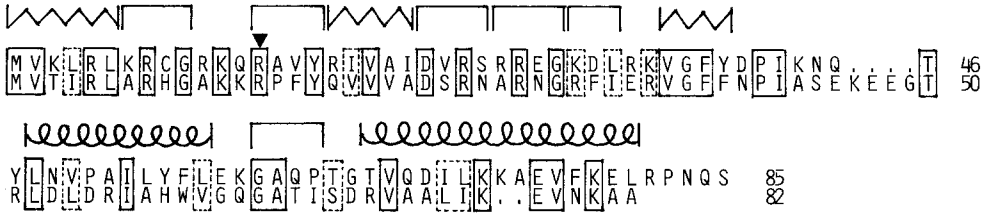

Fig. 3. Comparison between the amino acid sequence deduced from $ORF_{14}+ORF_{71}$ (CS16) and that of *Escherichia coli* S16 (ES16). Identical amino acid residues are in *solid boxes*. *Boxes with broken lines* indicate conservative replacements (D/E, K/R, S/T, L/I/M/V). Common secondary structures, alpha helix (*coils*), beta sheet (⋀⋀) and reverse turn (⌐¬), are shown over the amino acid sequences. A *triangle* indicates the intron site in *rps*16

1983). Complex secondary structures may be necessary for excising the group II introns.

Group III introns (Sugita et al. 1985; this paper). Introns of *trnG* (Deno and Sugiura 1984; Quigley and Weil 1985), *trnK* (Sugita et al. 1985), *rpl2* (Zurawski et al. 1984), *rps12* (Hallick et al. 1985), *atpF* (Bird et al. 1985) and *rps16* (this paper) have conserved sequences at their boundaries (Fig. 4). If the intron site of *trnV*, the gene for $tRNA^{Val}$ (UAC) (Deno et al. 1982; Krebbers et al. 1984; Zurawski and Clegg 1984), is shifted one base (to position 38/39) from the position suggested (Deno et al. 1982), its possible intron boundary sequence aligns well with the above sequences. The conserved boundary sequences of the group III introns are GTGCGNY at the 5′ ends and ATCN-RYY(N)YYAY at the 3′ ends (Fig. 4). These conserved

			5′EXON	INTRON	3′EXON	References
I	Tobacco	trnL-UAA	ACGGACTT	AATTGGATTGAGCCTTG--------GAAATTTATCGTAAGAGG	AAAATCCG	unpublished data
	Maize	trnL-UAA	ACGGACTT	GATTGTATTGAGCCTTG--------GAAATTTCTAGTAAAAGG	AAAATCCG	Steinmetz et al, 1982
	Bean	trnL-UAA	ACGGACTT	AATTGTATTGAGCCTTG--------GAAAGTTTTAGTAATCGG	AAAATCCG	Bonnard et al., 1983
II	Tobacco	trnI-GAU	CCTGATAA	TTGCGTCGTTGTGCCTG--------TGCGATGATTTACTTCAC	GGGGCGAG	Takaiwa & Sugiura, 1982
	Maize	trnI-GAU	CCTGATAA	TT CGTCGTTGTGCCTC--------GGCGATGATTTACTTCAC	GGGGCGAG	Koch et al., 1981
	Tobacco	trnA-UGC	TCTTGCAA	TTGGGTCGTTGCGATTA--------GGCGGTGGTTTACCCTGC	GGCGGATG	Takaiwa & Sugiura, 1982
	Maize	trnA-UGC	TCTTGCAA	TTGGGTCGTTGCGATTA--------GGCGGTGGTTTACCCTGT	GGCGGATG	Koch et al., 1981
III	Tobacco	trnV-UAC	GTTTACAC	GCGCGCCAATGTTTTTC--------AGTTTGACCTGTT TTAC	CGAGAAGG	Deno et al., 1982
	Maize	trnV-UAC	GTTTACAC	GTGCGCCAATGCTTTTC--------TGTTTGATCTGTT TTAC	CGAGAAGG	Krebbers et al., 1984
	Barley	trnV-UAC	GTTTACAC	GTGCGCCAATGCTTTTC--------CGTTTGATCGTTT TTAC	CGAGAAGG	Zurawski & Clegg, 1984
	Tobacco	trnG-UCU	TGGTAAAA	GTGTGATTCGTTCTATT--------ATCGTCGTCGACTATAAC	CCCTAGCC	Deno & Sugiura, 1984
	Wheat	trnG-UCU	TGGTAAAA	GTGTGATTCGTTCTATT--------ACCAACGTCGACTATAAC	CCCTAGCC	Quigley & Weil, 1985
	Tobacco	trnK-UUU	GCTTTTAA	GTGCGGCTAGTCTCTTT--------GAACTTATCTACT CCAT	CCGACTAG	Sugita et al., 1985
	Tobacco	rpl2	ACCTTTGA	GTGCGGTTTGAACTATT--------AGAAGAATCTACT TCAA	CCGATATG	Zurawski et al., 1984
	Tobacco	rps16	AAAGCAAC	GTGCGACTTGAAGGACA--------ATCTACATCAATCCCAAT	GAGCCGTC	this paper
	Tobacco	rps12	GCGTTCTA	GTGCGTTGTAGATT--------------ATGATCCACC CTAC	AATATGGG	Hallick et al., 1985
	Wheat	atpF	GGGAGTGT	GTGCGAGTTGTCTATTT--------CAAATAATCTACTTTCAT	TAAAAGAT	Bird et al., 1985
	Tobacco	atpF	GGGAGTGT	GTGCGAGTTGTTTATTT--------AAAATAATCTACTTTCAT	TAAGTGAT	unpublished data
Group III conserved sequence				GTGCGNY c t g	ATCNRYYNYYAY gc t c a a	this paper
Euglena conserved sequence				GTGYG t c	TARTTNTAY ct ga a a	Hallick et al., 1985
Nuclear gene conserved sequence			AAG c	GTGRAGT	YYYYYYNCAG GG t	Cech, 1983

Fig. 4. Comparison of exon-intron boundary sequences of chloroplast genes in higher plants. The conserved intron boundary sequences of *Euglena* chloroplast genes (Hallick et al. 1985) and of nuclear protein genes (Cech 1983) are also shown

boundary sequences resemble those of *Euglena gracilis* chloroplast introns (see Hallick et al. 1985). Moreover, they resemble the conserved boundary sequences of nuclear gene introns (see Cech 1983). The Group III introns may be spliced out by a similar mechanism to that which worked for nuclear mRNA precursors. Recently, Keller and Michel (1985) have reported that the introns of *trnV* and *trnG* belong to group II. However, both the 5′ and 3′ boundary sequences of these introns are quite different from those of *trnI* and *trnA*, which belong to group II. On the basis of the boundary sequences of introns, we think that the chloroplast introns can be classified into at least three groups and that *trnV* and *trnG* belong to group III.

Expression of rps16

Northern blot hybridization was carried out to examine whether tobacco *rps16* containing a long intron is expressed in chloroplasts. Total tobacco chloroplast RNA was electrophoresed in a 1.2% agarose gel. The RNA was transferred to nylon membrane sheets and hybridized with the 77 bp *Hpa*II-*Bam*HI fragment (positions −19 to 58) containing the 5′ exon, the 448 bp *Bam*HI + 282 bp *Bam*HI-*Eco*RI fragments (positions 59 to 789) containing the intron, and the 480 bp *Eco*RI fragment (positions 790 to 1,269) containing the 3′ exon. These three DNA fragments hybridized to an RNA band of about 1.3 kb (Fig. 5). From its size, the 1.3 kb RNA should be an unspliced precursor RNA. The DNA fragments containing the 5′ and 3′ exon also hybridized to 0.4 kb RNA, indicating that the 1.3 kb precursor is processed to the mature 0.4 kb RNA. The DNA fragments containing the intron also hybridized to 0.9 kb RNA which should be the spliced-out intron sequence. The DNA fragments containing the intron and 3′ exon hybridized to additional 0.7 and 0.5 kb RNAs. These RNAs may be processing intermediates. It is interesting that the ratio of 0.4 kb to 1.3 kb RNA is higher in the dark than under light (Fig. 5). The rate of transcription

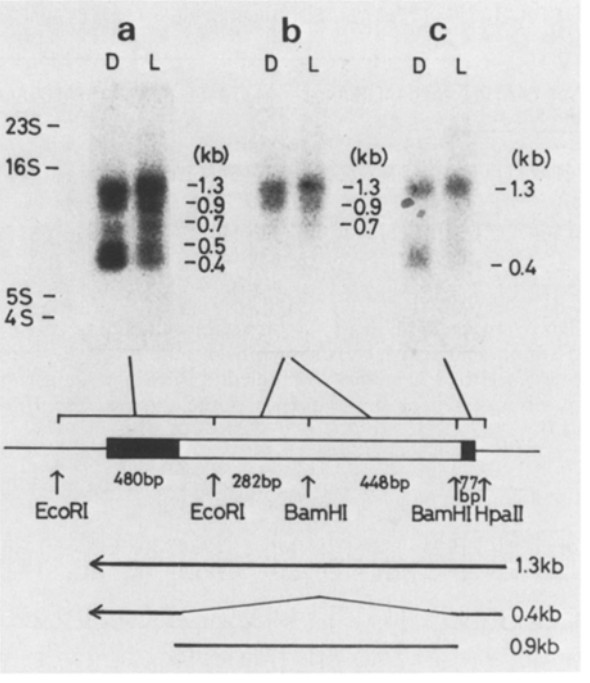

Fig. 5. Detection of precursor and mature mRNA molecules for CS16. Autoradiographs of RNA blots of tobacco chloroplast RNA hybridized with ^{32}P-labelled 480 bp *Eco*RI fragment (lane a), 448 bp *Bam*HI + 282 bp *Bam*HI–*Eco*RI fragments (lane b) and 77 bp *Hpa*II–*Bam*HI fragment (lane c). RNA samples were extracted from leaves grown under light (*L*) or in the dark (*D*). Size markers are *Escherichia coli* 23S, 16S, 5S and 4S RNAs. The *lower part* indicates the location of the DNA probes used. Exons are shown as *filled boxes* and an intron as an *open box*. *Arrows* indicate the major RNA species detected

is generally expected to be higher under light. It is suggested that splicing is the rate-limiting step in *rps16* expression.

The "Pribnow box"-like sequences TATGGT positions −85 to −80) and CATAAT (positions −41 to −36), and

the "−35 region"-like sequences TAGATA (positions −108 to −103) and ATGACT (positions −58 to −53) were found in the region upstream of the 5′ exon (Fig. 2). Two inverted repeated sequences (positions 1,153 to 1,198) were found in the region downstream of the 3′ exon. These structures may be transcriptional initiation and termination signals for *rps16*. The distance between these two structures is about 1.3 kb, which agrees well with the size of an *rps16* transcript detected by Northern blot hybridization.

Acknowledgements. We thank Prof. Mitsuru Takanami for advising us on the dideoxy chain termination method and Mr. Tohru Matsubayashi for the preparation of tobacco chloroplast RNA. This work was supported in part by Grants-In-Aid from the Ministry of Eduction, Science and Culture, and from the Toray Science Foundation.

References

Bartsch M, Kimura M, Subramanian AR (1982) Purification, primary structure, and homology relationships of a chloroplast ribosomal protein. Proc Natl Acad Sci USA 79:6871–6875

Bird CR, Koller B, Auffret AD, Huttly AK, Howe CJ, Dyer TA, Gray JC (1985) The wheat chloroplast gene for CF_0 subunit of ATP synthase contains a long intron. EMBO J 4:1381–1388

Bonnard G, Michel F, Weil JH, Steinmetz A (1984) Nucleotide sequence of the split $tRNA^{Leu}$-UAA gene from *Vicia faba* chloroplast. Mol Gen Genet 194:330–336

Cech T (1983) RNA splicing: three themes with variations. Cell 34:713–716

Chou PY, Fasman G (1978) Prediction of protein secondary structure. Adv Enzymology 47:45–148

Deno H, Kato A, Shinozaki K, Sugiura M (1982) Nucleotide sequence of tobacco chloroplast genes for elongator $tRNA^{Met}$ and $tRNA^{Val}$ (UAC): the $tRNA^{Val}$ (UAC) gene contains a long intron. Nucleic Acids Res 10:7511–7520

Deno H, Sugiura M (1983) The nucleotide sequences of $tRNA^{Ser}$ (GCU) and $tRNA^{Gln}$ (UUG) genes from tobacco chloroplasts. Nucleic Acids Res 11:8407–8414

Deno H, Sugiura M (1984) Chloroplast $tRNA^{Gly}$ gene contains a long intron in the D stem: Nucleotide sequences of tobacco chloroplast genes for $tRNA^{Gly}$ (UCC) and $tRNA^{Arg}$ (UCU). Proc Natl Acad Sci USA 81:405–408

Deno H, Shinozaki K, Sugiura M (1983) Nucleotide sequence of tobacco chloroplast gene for the α subunit of proton-translocating ATPase. Nucleic Acids Res 11:2185–2191

Deno H, Shinozaki K, Sugiura M (1984) Structure and transcription of a tobacco chloroplast gene for the III subunit of proton-translocating ATPase. Gene 32:195–201

Eneas-Filho JN, Hartley MR, Mache R (1981) Pea chloroplast ribosomal proteins: characterization and site of synthesis. Mol Gen Genet 184:484–488

Erickson JM, Rahire M, Rochaix J-D (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. EMBO J 3:2753–2762

Hallick RB, Gingrich JC, Johanningmeier U, Passavant C (1985) Introns in *Euglena* and *Nicotiana* chloroplast protein genes. In: Vloten-Doting L, Groot GSP, Hall TC (eds) Molecular form and function of the plant genome. Plenum Publishing Corp, New York, pp 211–220

Keller M, Michel F (1985) The introns of the *Euglena gracilis* chloroplast gene which codes for the 32-KDa protein of photosystem II. FEBS Lett 179:69–73

Koch W, Edwards K, Kossel H (1981) Sequence of the 16S–23S spacer in a ribosomal RNA operon of *Zea mays* chloroplast DNA reveals two split tRNA genes. Cell 25:203–213

Krebbers ET, Steinmetz A, Bogorad L (1984) DNA sequences for the *Zea mays* tRNA genes tV-UAC and ts-UGA: tv-UAC contains a large intron. Plant Mol Biol 3:13–20

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Messing J, Crea R, Seeburg PH (1981) A system for shotgun DNA sequencing. Nucleic Acids Res 9:309–321

Michel F, Dujon B (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. EMBO J 2:33–38

Montandon PE, Stutz E (1984) The genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*. Nucleic Acids Res 12:2851–2859

Ohme M, Kamogashira T, Shinozaki K, Sugiura M (1985) Structure and cotranscription of tobacco chloroplast genes for $tRNA^{Glu}$ (UUC), $tRNA^{Tyr}$ (GUA) and $tRNA^{Asp}$ (GUC). Nucleic Acids Res 13:1045–1056

Quigley F, Weil JH (1985) Organization and sequence of five tRNA genes and of an unidentified reading frame in wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr Genet 9:495–503

Rochaix J-D, Rahire M, Michel M (1985) The chloroplast ribosomal intron of *Chlamydomonas reinhardii* for a polypeptide related to mitochondrial maturases. Nucleic Acids Res 13:975–984

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Shinozaki K, Deno H, Kato A, Sugiura M (1983) Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase. Gene 24:147–155

Steinmetz A, Gubbins EJ, Bogorad L (1982) The anticodon of the maize chloroplast gene for $tRNA^{Leu}$ (UAA) is split by a large intron. Nucleic Acids Res 10:3027–3037

Subramanian AR, Steinmetz A, Bogorad L (1983) Maize chloroplast DNA encodes a protein sequence homologous to the bacterial ribosome assembly protein S4. Nucleic Acids Res 11:5277–5286

Sugita M, Sugiura M (1983) A putative gene of tobacco chloroplast coding for ribosomal protein similar to *E. coli* ribosomal protein S19. Nucleic Acids Res 11:1913–1918

Sugita M, Shinozaki K, Sugiura M (1985) Tobacco chloroplast $tRNA^{Lys}$ (UUU) gene contains a 2.5 kilobase pair intron: An open reading frame and a conserved boundary sequence in the intron. Proc Natl Acad Sci USA 82:3557–3561

Sugiura M, Kusuda J (1979) Molecular cloning of tobacco chloroplast ribosomal RNA genes. Mol Gen Genet 172:137–141

Takaiwa F, Sugiura M (1982) Nucleotide sequence of the 16S-23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucleic Acids Res 10:2665–2676

Umesono K, Inokuchi H, Ohyama K, Ozeki H (1984) Nucleotide sequence of *Marchantia polymorpha* chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of *E. coli* ribosomal protein S14. Nucleic Acids Res 12:9551–9565

Vandekerckhove J, Rombauts J, Wittmann-Liebold B (1977) The complete amino acid sequence of protein S16 from *Escherichia coli*. Hoppe Seylers Z Physiol Chem 358:989–1002

Whitfeld P, Bottomley W (1983) Organization and structure of chloroplast genes. Annu Rev Plant Physiol 34:279–310

Zurawski G, Clegg MT (1984) The barley chloroplast DNA *atp*BE, *trnM2 and trn*V1 loci. Nucleic Acids Res 12:2549–2559

Zurawski G, Bottomley W, Whitfeld PR (1984) Junctions of the large single copy region and the inverted repeats in *Spinacia oleracea* and *Nicotiana debneyi* chloroplast DNA: sequence of the genes for $tRNA^{His}$ and the ribosomal proteins S19 and L2. Nucleic Acids Res 12:6547–6558

Communicated by K. Isono

Received July 25 / September 6, 1985