Volume 14 Number 9 1986 Nucleic Acids Research

Peculiar feature of the organization of rRNA genes of the *Chlorella* chloroplast DNA

Takashi Yamada and Miyuki Shimaji

Department of Cell Biology, Mitsubishi-Kasei Institute of Life Sciences, 11 Minamiooya, Machida-shi, Tokyo 194, Japan

Received 7 February 1986; Revised and Accepted 4 April 1986

ABSTRACT

The organization of a cloned rRNA gene cluster from Chlorella ellipsoidea chloroplast DNA (cpDNA) has been analyzed. Southern hybridization experiments with labelled chloroplast rRNAs as probes revealed an extraordinarily large size of the 16S-23S rRNA spacer region, ca. 4.8 kbp, almost twice as large as those of most higher plants. The nucleotide sequence determined on this region has shown that: (1) The $tRNA^{Ile}$ gene locating in this region is similar to those of higher plant chloroplasts, blue-green algae and E. coli but does not contain any introns in contrast to higher plant chloroplasts. (2) The $tRNA^{Ala}$ gene is absent from this region. (3) There are four open reading frames (ORFs) coding for 55, 102, 107 and 110 amino acids, respectively. (4) A few sets of unique sequence were found repeatedly in this region. (5) The 23S rRNA gene is coded on the opposite strand in the reverse order. This arrangement of the 16S-23S rRNA region of Chlorella cpDNA is quite different from any of those reported so far for various organisms.

INTRODUCTION

One of the outstanding differences in the organization of rRNA operon between higher plant chloroplasts and prokaryotes such as blue-green algae and E. coli is in the size of the 16S-23S rRNA spacer region. In general, the spacer region of cpDNA from most higher plants is considerably larger (ca. 2 kbp) than that of prokaryotic genomic rrn operons (ca. 500-600 bp) (1). This is attributable mainly to, as demonstrated in maize (2) and tobacco (3), the presence of large introns (ca. 700-950 bp) in the genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ located within the spacer region of higher plants.

The corresponding region of algal cpDNAs so far characterized, although only a few organisms have been studied, is somewhat different from that of higher plants. In Euglena gracilis chloroplasts, the spacer region is rather small (259 bp) and the two tRNA genes do not contain any introns (4). The chloroplast rRNA operon of the unicellular green alga Chlamydomonas reinhardii has a larger 16S-23S

spacer region (ca. 1.7 kbp) where two additional small 7S and 3S rRNAs are located (5) and several short direct and inverted repeat sequences are found (6). The $tRNA^{Ile}$ and $tRNA^{Ala}$ genes located within this spacer do not contain introns (Schneider and Rochaix, unpublished results). Thus, the organization of chloroplast rRNA operons seems to be considerably variable among algal speices. In other words, the well-documented features of chloroplast genes that are thought to be conserved among higher plants (1) might not be necessarily shared within the algal world.

We have been interested in the molecular arrangement of cpDNA from the unicellular green alga Chlorella ellipsoidea. This cpDNA is a ca. 175 kbp circular molecule containing a pair of inverted repeat of 22.5 kbp in size, where two sets of the rRNA genes are located (7). In this report, we cloned a whole set of chloroplast rRNA genes from C. ellipsoidea and determined the nucleotide sequence of its 16S-23S rRNA region with special reference to its spacer. A peculiar feature of the organization of this region was found.

MATERIALS AND METHODS

Strains. Chlorella ellipsoidea C-87 was obtained from the algal culture collection of the Institute of Applied Microbiology, Univ. of Tokyo. Escherichia coli HB101, JM101 and JM109 (8) were used for bacterial transformation and propagation of plamsids.

DNAs and RNAs. Chloroplast DNA from C. ellipsoidea was prepared as described previously (9). Plasmid DNAs were prepared according to Maniatis et al. (10). A gene library of C. ellipsoidea cpDNA was constructed by ligation of SstI-digested cpDNA fragments and SstI-digested pUC13 (11) and transformed into E. coli HB101 and JM109. Ribosomal RNAs were prepared as described preivously (7).

Gel Electrophoresis and Southern Hybridization. Chloroplast DNA and plasmid DNAs digested with restriciton enzymes were analyzed by electrophoresis on horizontal agarose gels. Digestion of DNA with various restriction enzymes (Takara Shuzo Co., Kyoto; Toyobo Biochemicals, Osaka; Bethesda Research Labs, Maryland; New England Biolabs, Boston) was performed according to the suppliers. For hybridization analysis, electrophoretically separated DNA was transferred to a nitrocellulose filter and hybridized with ^{32}P-

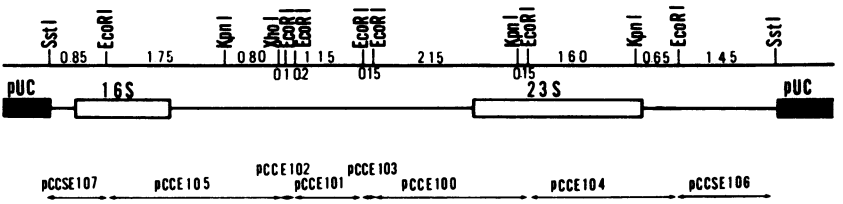


Figure 1. Physical map of the 11.0 kbp SstI-SstI insert of pCCS123. Subclones derived from this sequence are indicated by arrows. Sizes are shown in kilo base pairs.

labelled probes according to Southern (12). Ribosomal RNAs were labelled with [γ-^{32}P] ATP by T4 polynucleotide kinase (Takara Shuzo Co., Kyoto) reaction according to Sugiura and Kusuda (13).

Sequencing of DNA Fragments. Restriction fragments containing the Chlorella chloroplast rDNA were cloned in M13 mp18 and 19 (11). Single-stranded DNAs prepared from phage plaques bearing the recombinant DNAs were sequenced using the chain termination procedure (14). M13 sequencing kits were obtained from Takara Shuzo Co., Kyoto. S1 nuclease mapping was carried out as described by Berk and Sharp (15).

RESULTS

Cloning of the Chlorella Chloroplast rRNA Genes.

The rRNA genes of Chlorella cpDNA were previously shown to be located within a 11.0 kbp SstI fragment on the inverted repeat sequence (7). A clone containing this fragment was selected by colony hybridization with ^{32}P-labelled rRNA as probes from the Chlorella cpDNA gene library. A physical map of this clone (pCCS123) was consistent with the previous one (7) and is shown in Fig. 1.

The location of the 16S and 23S rRNA genes were determined by Southern hybridization using ^{32}P-labelled Chlorella chloroplast 16S and 23S rRNAs. Figure 2 shows the hybridization patterns: The 16S rRNA hybridized with fragments of 0.85 kbp EcoRI-SstI and 1.75 kbp EcoRI-KpnI, which indicates the 3' terminus of the 16S rRNA gene is within the 1.75 kbp EcoRI-KpnI fragment. The 23S rRNA hybridized with fragments of 0.65 kbp EcoRI-KpnI, 1.60 kbp KpnI-EcoRI and 2.15 kbp KpnI-EcoRI, indicating one terminus of the 23S rRNA gene is within the 2.15 kbp EcoRI-KpnI fragment.

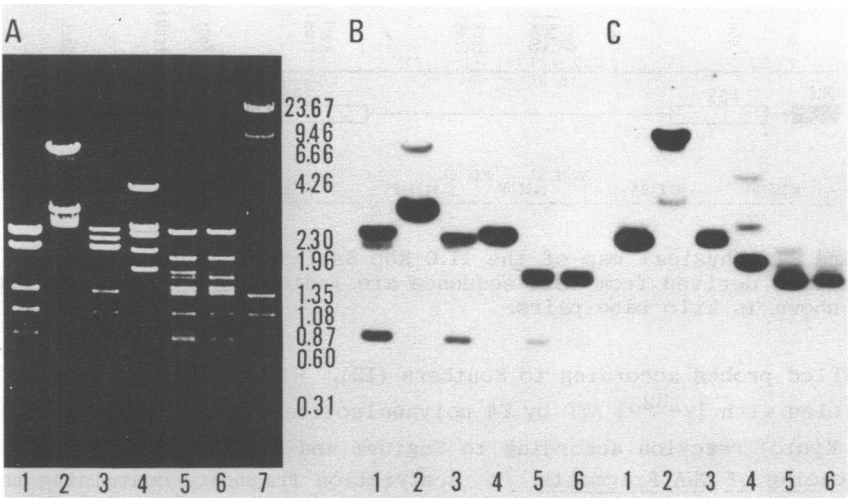


Figure 2. Hybridization of ^{32}P-labelled chloroplast rRNAs to the restriction fragments of pCCS123. (A) Agarose gel electrophoresis patterns of pCCS123 digested with EcoRI and SstI (lane 1); SstI and XhoI (lane 2); EcoRI, SstI and XhoI (lane 3); KpnI and SstI (lane 4); EcoRI, KpnI and SstI (lane 5); and EcoRI, KpnI, SstI and XhoI (lane 6). Lane 7 contained a mixture of λDNA digested with HindIII and ϕX174 RF DNA digested with HaeIII as size markers. (B) Hybridization patterns with 5'-labelled Chlorella chloroplast 16S rRNA. (C) Hybridization patterns with 5'-labelled Chlorella chloroplast 23S rRNA. Some additional faintly hybridizing bands are attributable to a cross-contamination of each rRNA. Sizes are shown in kilo base pairs.

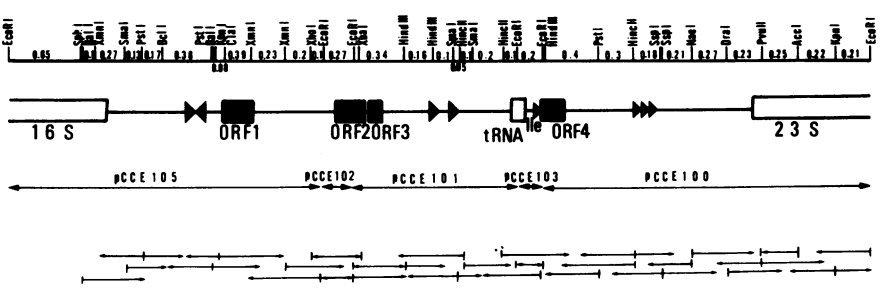


Figure 3. Physical map of the 16S-23S rRNA spacer region from Chlorella cpDNA and sequencing strategy. The lower part show the sequencing strategy. Horizontal arrows indicate the direction and extent of DNA regions sequenced. Sizes are shown in kilo base pairs.

For further analyses, the 11.0 kbp SstI insert of pCCS123 was digested with SstI and EcoRI into 8 fragments and all of them were ligated to pUC13 and transformed into E. coli JM109. The subclones obtained were designated as shown in Fig. 1.

Characterization of the 16S-23S Spacer Region.

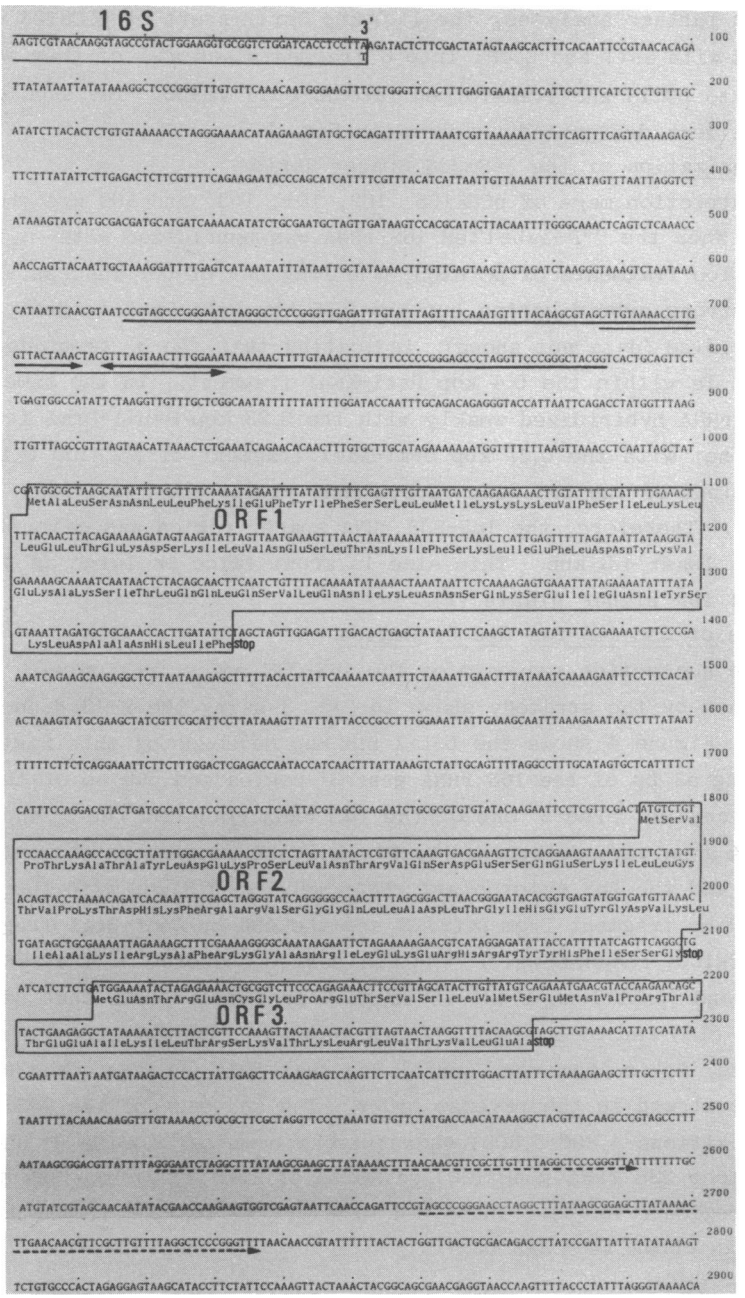
Restriction maps of pCCE105, 102, 101, 103, and 100 are shown in Fig. 3. When the ^{32}P-labelled 16S rRNA was hybridized with restriction fragments of pCCE105, the fragment of 0.4 kbp PstI-XmnI showed strong hybridization but the 0.55 kbp PstI-PstI fragment did not hybridize (data not shown), indicating that the 3' terminus of the 16S gene is within the 0.4 kbp PstI-XmnI fragment. In the same way, the 23S rRNA hybiridized weakly with the 0.23 kbp PvuII-DraI fragment but did not with the 0.27 kbp DraI-NaeI fragment of pCCE100 (Fig. 3); one terminus of the 23S gene is confined within the PvuII-DraI fragment. Therefore, the 16S-23S rRNA spacer region was calculated to be of at least 4.8 kbp. This size is about twice as large as that reported for higher plants (1).

Sequence of the 16S-23S Spacer Region.

The nucleotide sequence of the 16S-23S spacer region was determined by the strategy shown in Fig. 3 using the M-13 dideoxy method. Figure 4 shows the total 5.5 kbp sequence of this region including 52 bp of the 16S rRNA gene-3' region and 509 bp of the 23S rRNA 3' region. The 3' end of the 16S rRNA gene and the 3' end of the 23S rRNA gene were determined by S1 nuclease mapping experiments where the 16S and the 23S chloroplast rRNAs were hybridized with the XmnI-PstI 400 bp fragment from pCCE105 and the 406 bp AccI-DraI fragment from pCCE100, respectively. The nucleotide sequence of the 16S rRNA gene-3' end (positions 1-52) is in good accordance with that of tobacco (16) and of E. coli (17) as shown in Fig. 4. As for the 23S rRNA gene, it is, however, surprising that this gene is coded on the opposite strand in the reverse order. The sequence of the 23S rRNA gene (positions 4,896-5,500) shows strong homology to the 3' end of the tobacco 23S rRNA gene (18) in spite of lacking in the 23S-4.5S spacer region (Fig. 4). Thus, net size of the 16S-23S rRNA spacer of Chlorella cpDNA is 4,842 bp.

Gene for tRNAs.

The gene for tRNA^{Ile} was found at positions 3091-3164. In contrast to most higher plants, the Chlorella chloroplast tRNA^{Ile} gene


16 S
3'
ORF1
ORF2
ORF3
stop

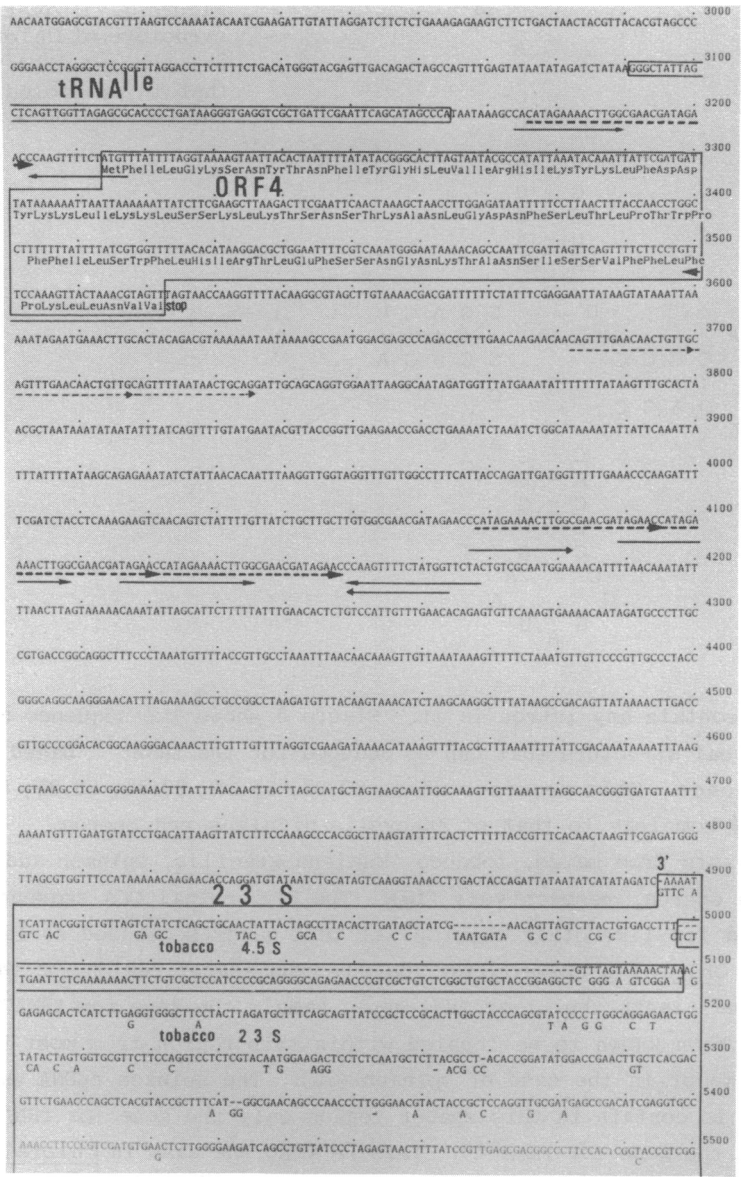


Figure 4. Nucleotide sequence of the 16S-23S rRNA spacer region from Chlorella cpDNA. The non-coding strand (RNA-like sequence) is shown. For comparison, the tobacco sequences for 16S and 23S rRNA genes are shown below the Chlorella sequence. Coding regions are boxed. Main tandem and inverted repeats are indicated by solid and dotted arrows, respectively.

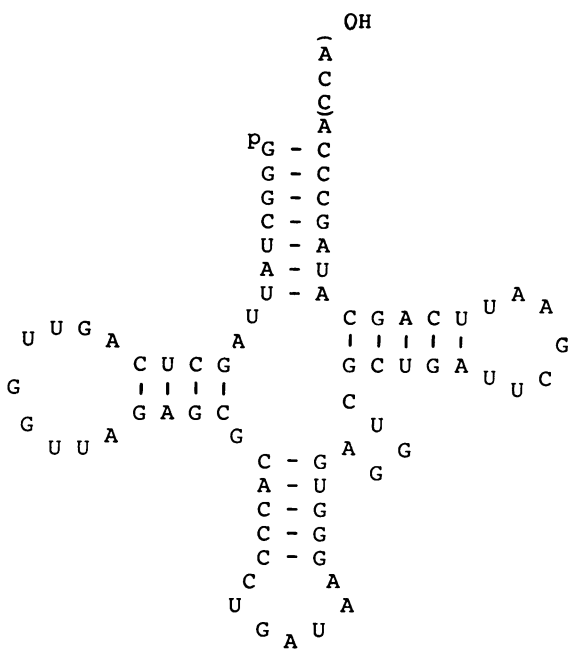


Figure 5. Clover leaf structure of Chlorella chloroplast $tRNA^{Ile}$ that is predicted from the DNA sequence.

did not contain any intron in it. Figure 5 shows the sequence and clover leaf structure that can be deduced for the $tRNA^{Ile}$ based on the DNA sequence. This sequence shows 89.2%, 85.1%, 83.8%, 83.8%, 54.1% and 79.7% homology to that of Anacystis nidulans rrn operon, chloroplasts from maize, tobacco, Euglena gracilis, spinach and E. coli rrn operon, respectively (19). The 3' terminal CCA sequence is not coded for like other chloroplast tRNA genes sequenced so far (1).

The Chlorella chloroplast 16S-23S rRNA spacer region, on the other hand, lacks the other tRNA gene, namely the gene for $tRNA^{Ala}$ that has been known to be located within this region from most higher plants except in the case of spinach (20). The spinach cpDNA was reported to contain in this spacer region only the gene for $tRNA^{Ile}$ and lack the gene for $tRNA^{Ala}$. This is also the case in Chlorella cpDNA.

Open Reading Frames.

Four open reading frames (ORF) are found in the 16S-23S rRNA spacer region, all of which are in the same direction as the 16S rRNA gene (Fig. 4). ORF1 (positions 1003-1333), ORF2 (positions 1793-2098), ORF3 (2112-2277) and ORF4 (3214-3523) could possibly code for

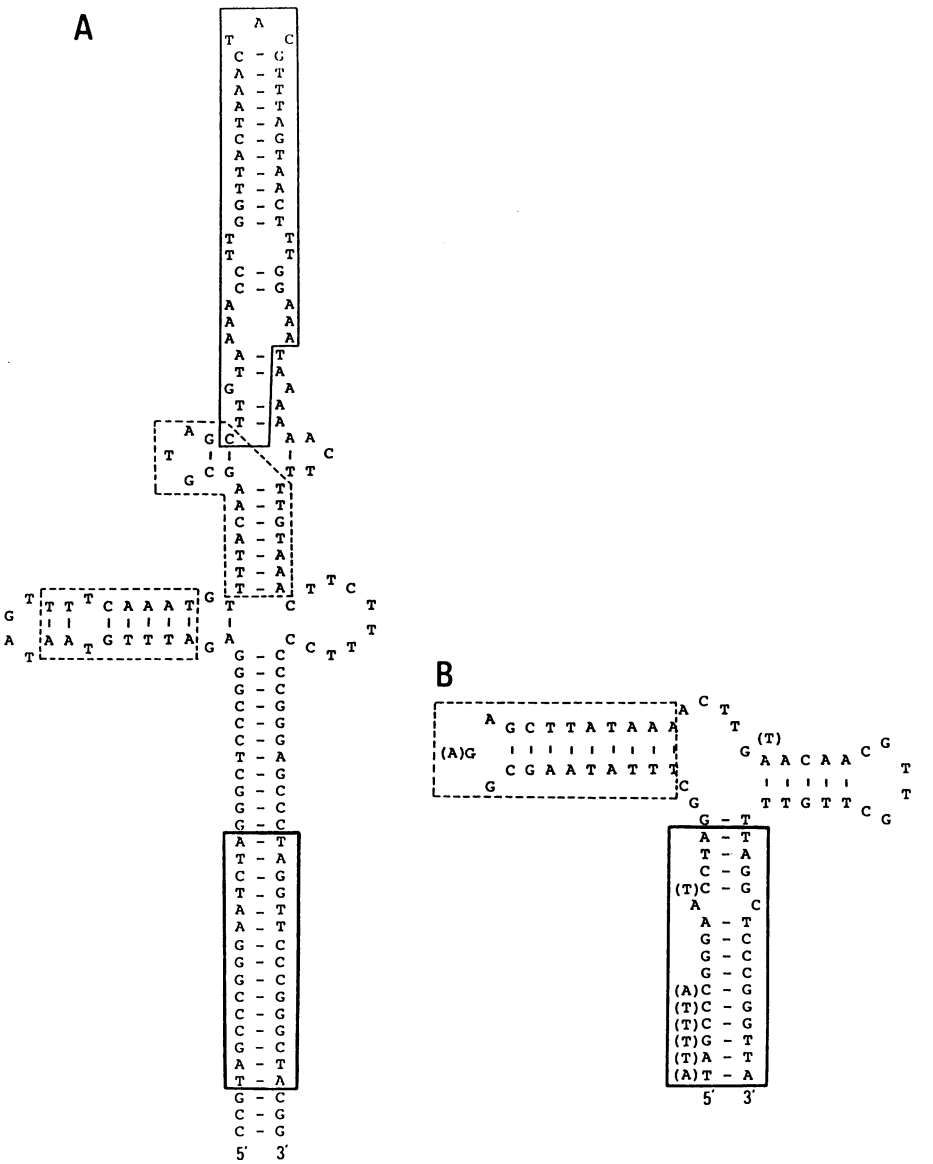


Figure 6. Possible secondary structures of the repetitive sequences flanking ORFs found in the 16S-23S rRNA spacer. (A) The inverted repeat sequence found at positions 617-787. (B) The tandem repeat sequence found at positions 2660-2736. The differences with another repeat (2521-2591) are indicated by parentheses. Thick solid box, dotted box and solid box indicate σ-elements, a conserved region among these repetitive sequences and a 45 bp sequence that is invertedly repeated at positions 3498-6543, respectively.

polypeptides of 110, 102, 55 and 107 amino acids, respectively. These ORFs show no appreciable homology to each other and to those found in the introns of $tRNA^{Ile}$ and $tRNA^{Ala}$ of the maize 16S-23S rRNA spacer region (2), in the Chlamydomonas reinhardii 23S rRNA intron (21), and in the small single copy side of the tobacco cpDNA inverted repeat (22). It is noteworthy that ORF4 located close to the gene for $tRNA^{Ile}$ shows about 20% amino acid sequece homology to one of two ORFs (ORF1, 98 a.a.) reported to be present between the genes for $tRNA^{Ser}$ and $tRNA^{Glu}$ from tobacco chloroplast (23). Transcripts of these ORFs in the Chlorella 16S-23S rRNA spacer have not been detected so far.

Inverted and Tandem Repeats.

The DNA sequence shown in Fig. 4 contains numerous direct and inverted repeats outside of the ribosomal coding regions. These repeats are composed of more than 10 contiguous bases and are located relatively close to each other. Their positions are indicated in Fig. 4. Above all, the most interesting sequences would be a pair of 84 bp-inverted repeats (positions 617-787) and two ca. 75 bp-tandem repeats (positions 2521-2591 and 2660-2736) which are sandwiching three of four ORFs (Fig. 4). These repeated sequences contain on their ends a conserved element consisting of ca. 15 bp (σ-element: TAGCCCGGGAACCTA, 5'-3'). Secondary structures of these inverted and tandem repeats are shown in Fig. 6. A 45 bp sequence flanked by σ-elements are invertedly repeated at the 3' end of ORF4 (positions 687-731 and 3,499-3,543).

The ORF4 is also flanked by a pair of 15 bp-inverted repeats (positions 3175-3217) and three 18 bp-tandem repeats (positions 3682-3735). It would be noteworthy to point out that the sequence of 27 nucleotides including the inverted repeats CATAGAAAACTTGGCGAACGATAGAAC (σ-element: positions 3176-3202) is tandemly repeated three times between ORF4 and the 23S rRNA gene (positions 4068-4148).

DISCUSSION

Organization of the 16S-23S rRNA Spacer Region from Chlorella Chloroplast DNA.

The spacer region of the Chlorella cpDNA 16S-23S is unusually large. This can be explained by the presence of additional two sets of repetitive sequence-flanked ORFs; ORF1-2-3 cluster flanked by σ-elements (positions 616-2,736; 2,126 bp) and ORF4 flanked by α-

elements (positions 3,175-4,148; 974 bp). These structures, especially the σ-element-associating structure remind us of not only the introns found in the genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ of higher plant rRNA spacer but also some kind of mobile genetic elements such as prokaryotic IS or Tn elements (24) and Ty elements in yeasts (25). Introns found in the genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ of maize contain each-related ORFs and a terminal inverted repeat structure (2), but those ORFs and repeated sequeces seem not to be directly relevant to those found in the Chlorella sequence. Any of the ORFs found in Chlorella have not so far been able to be expressed in vivo; this is also the case for ORFs in the tRNA introns.

Though there have been so far no clear-cut example of the transposition of cpDNA sequence in higher plants and algae, the peculiar structure of the Chlorella 16S-23S rRNA spacer suggests that extended rearrangements might have occurred through some insertion/deletion mechanisms via repetitive sequences. In this sense, it is noteworthy that Southern hybridization experiments using the ca. 200 bp HindIII-HindIII fragment that includes a part of σ-element-flanking repeat (Fig. 3, positions 2,543-2,766) as a probe showed the presence of some related sequences to the σ-elements on several other parts of cpDNA and even on the nuclear DNA from Chlorella (Yamada and Shimaji, unpublished reuslt). Recently, we revealed the presence of such a repetitive sequence with SmaI sites (like σ-element) in a Chlorella cpDNA fragment that shows ARS activity (26). Schneider et al. reported in the flanking regions of the chloropolast rRNA unit of Chlamydomonas reinhardii numerous tandem and iverted repeats (6). Since the 16S-23S rRNA spacer region of C. reinhardii is also relatively large and contains several repeated sequences (6), some sequences similar to σ- or α-elements and ORFs found in Chlorella might be present in C. reinhardii and also in other unicellular green algae.

Inversion of the gene for 23S rRNA.

Sequence analysis of the 16S-23S rRNA spacer revelaed another uncommon feature of organization, the inversion of the gene for 23S rRNA. Five independently isolated clones of the 11.0 kbp SstI fragment of the cpDNA consistently gave the same restriction maps as shown in Fig. 1 and Fig. 3. Furthermore, these maps are in good accordance with that made previously by restriction enzyme digestion

of the whole cpDNA and Southern hybridization with rDNAs (7). Therefore, the inversion could not be occurred artificially through cloning steps. The whole sequence of the 23S rRNA gene has shown strong homology to that of tobacco except lacking in the 23S-4.5S rRNA spacer and carrying a 249 bp intron (Yamada and Shimaji, in preparation). Thus, this gene would not be an apparent pseudogene.

Two questions arise here; (1) by what mechanism this reversion occurred and (2) from which promoter this gene is transcribed. Palmer pointed out that inversions observed in several higher plant cpDNAs have one endpoint just downstream from the gene for atpA (27). In tobacco, a 3.5 kbp region immediately downstream from atpA has been sequenced and found to contain four tRNA genes and two ORFs (23). It is intriguing that one of the two ORFs of that region has a ca. 20% sequence homology to ORF4 found in this study. Furthermore, in Chlorella, the gene for $tRNA^{Ala}$ is lost, whereas the gene for $tRNA^{Ile}$ is present intactly; one endpoint of inversion seem to be confined within the tRNA spacer region. Since that tobacco region contains four tRNA genes, tRNA sequences may somehow have to do with the inversion. Alternatively, this inversion may have occurred within the inverted repeat sequence on the small single copy side, where the $tRNA^{Asn}$ gene and its promoter sequence is known to be located on the opposite strand in a reverse direction to the 16S rRNA transcription in the tobacco cpDNA (22). If this is the case, the 23S rRNA gene must be transcribed from the promoter of the $tRNA^{Asn}$ gene in a reverse direction. The fine structure of the upstream region of the 23S rRNA gene and its initiation site of transcription is remaining to be characterized.

In this paper, we described an unusual organization of rRNA operons from Chlorella cpDNA. This example shows that the highly conserved chloroplast gene arrangement of higher plants may not always fit in the algal world, where a considerably greater heterogeneity in both size and structure of cpDNAs is known (27). In order to understand the whole figure of chloroplast gene arrangement, algal chloroplast genes should be studied more extensively.

ACKNOWLEDGEMENTS

We are very grateful to Drs Makoto Kageyama, Kenji Sakaguchi and

Hiroshi Mizutani for helpful discussions and critical reading of the manuscript. We also thank Miss Tomoko Mukohara for typing the manuscript.

REFERENCES

1. Whitfeld, P.R. and Bottomley, W. (1983) Ann. Rev. Plant Physiol. 34, 279-310.
2. Koch, W., Edwards, K. and Kössel, H. (1981) Cell 25, 203-213.
3. Takaiwa, F. and Sugiura, M. (1982) Nucl. Acids Res. 10, 2665-2676.
4. Graf, L. Kössel, H. and Stutz, E. (1980) Nature 286, 908-910.
5. Rochaix, J. D. and Darlix, J.L. (1982) J. Mol. Biol. 159, 383-395.
6. Schneider, M., Darlix, J.L. and Rochaix, J.D. (1985) Nucl. Acids Res. 23, 8531-8542.
7. Yamada, T. (1983) Current Genet. 7, 481-487.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
9. Yamada, T. (1982) Plant Physiol. 70, 92-96.
10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in Molecular Cloning, A Laboratory Manual, pp.86-369, Cold Spring Harbor Lab. Press, New York.
11. Horrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101-106.
12. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
13. Sugiura, M. and Kusuda, J. (1979) Mol. Gen. Genet. 172, 137-141.
14. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA. 74, 5463-5467.
15. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
16. Tohdoh, N. and Sugiura, M. (1982) Gene 17, 213-218.
17. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801-4805.
18. Takaiwa, F. and Sugiura, M. (1982) Eur. J. Biochem. 124, 13-19.
19. Sprinzl, M., Vorderwülbecke, T. and Hartmann, T. (1985) Nucl. Acids Res. 13, Supplement r51-r104.
20. Bohnert, H.J., Driesel, A.J., Crouse, E.J., Gordon, K. and Herrmann, R.G. (1979) FEBS Lett. 103, 52-56.
21. Rochaix, J.D. Rahire, M. and Michel, F. (1985) Nucl. Acids Res. 13, 975-984.
22. Sugita, M., Kato, A., Shimada, H. and Sugiura, M. (1984) Mol. Gen. Genet. 194, 200-205.
23. Deno, H. and Sugiura, M. (1983) Nucl. Acids Res. 11, 8407-8414.
24. Iida, S., Meyer, J. and Arber, W. (1983) in Mobile Genetic Elements (Shapiro, J.A. ed.) Academic Press Inc., New York, pp.159-222.
25. Roeder, G.S. and Fink, G.R. (1983) in Mobile Genetic Elements (Shapiro, J.A. ed) Academic Press Inc., New York, pp. 300-328.
26. Yamada, T., Shimaji, M. and Fukuda, Y. (1986) Plant Mol.Biol. 6, in press.
27. Palmer, J.D. (1985) Ann. Rev. Genet. 19, 325-354.