Volume 12 Number 24 1984 Nucleic Acids Research

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

Kazuhiko Umesono[1], Hachiro Inokuchi[1], Kanji Ohyama[2] and Haruo Ozeki[1]*

[1]Department of Biophysics, Faculty of Science, and [2]Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan

Received 16 July 1984; Revised and Accepted 7 November 1984

ABSTRACT

The nucleotide sequence of a region of Marchantia polymorpha chloroplast DNA was determined. On this DNA sequence (3.38kb), three open reading frames (ORFs) and three putative tRNA genes were detected in the following order: -ORF701-tRNA^{Ser}(UGA)-ORF702-tRNA^{Gly}(GCC)-initiator tRNA^{Met}(CAU)-ORF703-. The ORF703 is composed of 100 codons in which those for lysine (15%) and arginine (11%) are abundant, and could be accounted for as a counterpart of E. coli ribosomal protein S14 since they share 45% homology in the amino acid sequences. The ORF701 appears to code for a membrane protein, showing a periodic appearance of seven clusters of hydrophobic amino acids. Although the mechanisms remain unknown, the ORF701 causes a streptomycin-sensitive phenotype in resistant mutants of E. coli. The ORFs and tRNA genes are separated from each other by extremely AT-rich spacers containing sequences of dyad symmetry. The third letter positions of the codons in the ORFs are also rich in A and T residues.

INTRODUCTION

A number of works on the structure and organization of chloroplast genes have been reported, but they are mostly on those of higher plants or algae. Among the lower plants, a physical map of chloroplast DNA has been so far constructed only in the case of liverwort (1). Although the estimated size of chloroplast DNA from the liverwort Marchantia polymorpha L. (ca. 120kb) is somewhat smaller than those from higher plants (ca. 150kb (2)), the gross organization of many chloroplast genes in the former DNA including the relative positions of the two invertedly oriented rRNA genes and the LS gene (for the large subunit of RuBP carboxylase) appears to be similar to that of the corresponding genes in the latter chloroplast DNAs (1). A physical map of the entire chloroplast DNA of M. polymorpha has been constructed with the restriction endonucleases BamHI, SmaI, KpnI, and XhoI (1). The Bam7 DNA fragment (4.9kb BamHI fragment), which is analyzed herein in detail, is located almost at the center of the large single-copy region of M. polymorpha chloroplast DNA. In the center of this region, several relatively small BamHI fragments (less than 5kb) are clustered in the order "-Bam13-Bam7-Bam8-Bam10-

..(LS).." (1). In this order, a minor correction has been made recently in the course of cloning and other experiments, that the 7-8 region is in reverse orientation with respect to the Bam13 and Bam10 fragments; thus, the newly assigned order is -Bam13-Bam8-Bam7-Bam10- (in preparation, Y. Yamashita, K. Ohyama, and T. Komano). In our present experiments, the Bam7 fragment of M. polymorpha chloroplast DNA was cloned at the BamHI site of pBR322, and the nucleotide sequence (3.38kb) starting from the BamHI site at the junction of the Bam8 and Bam7 fragments was determined. On this DNA sequence, we were able to detect three putative tRNA genes and three possible open reading frames. We report here the structure and organization of these putative genes in this region of the liverwort chloroplast DNA.

MATERIALS AND METHODS

Recombinant plasmids carrying liverwort chloroplast DNA fragments

Chloroplast DNA of M. polymorpha was isolated from suspension cultures of green cells in covalently closed circular (ccc) form (3). The DNA was digested completely with BamHI and the fragments were cloned at the BamHI site of pBR322. One of the recombinant plasmids, named pMP227, was used as a sole material which carries a 4.9kb BamHI fragment. This fragment has been assigned to the Bam7 fragment of M. polymorpha chloroplast DNA (1).

Bacterial strains and culture media

The following strains were used, all of which were streptomycin (Sm) resistant mutants (strA) of E. coli; HB101 (4), NO1345 (5) and BT52recA (lac_{amb}, trp_{amb}, recA, strA. this laboratory). For plasmid isolation, bacteria were grown in E-broth (6). The Sm plate used was lambda agar (7) containing streptomycin sulfate (Meiji Co. Ltd.).

DNA sequencing

The plasmid pMP227 was propagated in E. coli HB101. The plasmid DNA in ccc form was purified by EtBr/CsCl centrifugation. After digestion with an appropriate restriction enzyme, the DNA fragments were separated by agarose gel electrophoresis. For DNA sequencing, the fragments were extracted from the gel and treated with calf intestinal alkaline phosphatase (Boeringer Mannheim GmbH). Labeling at 5' ends was performed with T4 polynucleotide kinase (Takara Syuzo Co. Ltd.). DNA sequencing was carried out acccording to the method of Maxam and Gilbert (8).

RESULTS AND DISCUSSION

Structure and organization of putative genes in cloned Bam7 fragment

The 4.9kb BamHI fragment cloned in pMP227 electrophoretically comigrated

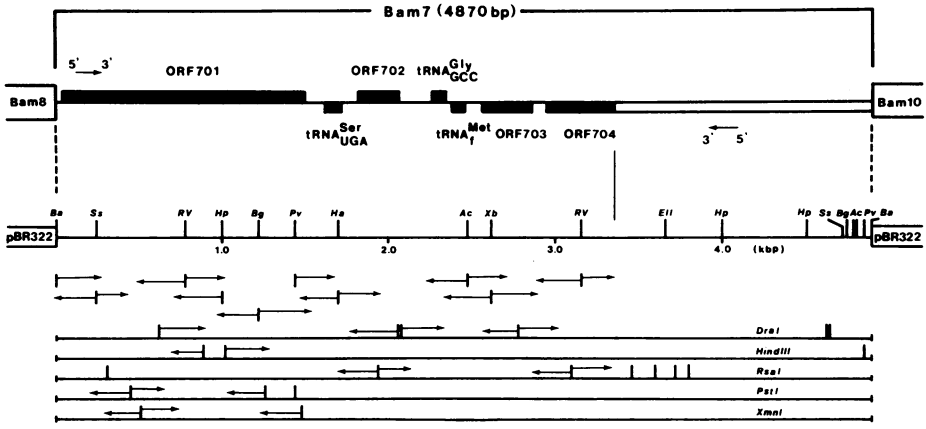


Fig. 1 Physical map and sequencing strategy of the cloned Bam7 fragment. The upper figure shows the orientation of the Bam7 fragment on the liverwort chloroplast DNA in relation to the Bam8 and Bam10 fragments, and the order as well as the direction of the genes were deduced from the DNA sequence (Fig. 2). In the lower figure, the recognition sites of restriction enzymes used in the DNA sequencing are mapped. Arrows indicate the direction of the sequencing. Ac,AccI; Ba,BamHI; Bg,BglII; EII,BstEII; Ha,HaeIII; Hp,HpaII; Pv,PvuII; RV,EcoRV; Ss,SstI; Xb,XbaI.

with Bam7 of M. polymorpha chloroplast DNA (data not shown). A physical map of the Bam7 fragment and the sequencing strategy used in the study are shown in Fig. 1. The DNA sequence of part of the Bam7 fragment (3,381bp from one of the BamHI ends) is presented in Fig. 2.

With the aid of a computer, three possible open reading frames and three tRNA-like sequences were detected in the region sequenced. We called these open reading frames ORF701 (positions 44-1504 for 487 codons; calculated molecular weight of the predicted protein, 53,430 daltons), ORF702 (1824-2054 for 77 codons; 8,520 daltons), and ORF703 (2558-2857 for 100 codons, 11,880 daltons). The sequence from position 2949 to the end (3381) also appears to represent a distal portion of a presumed open reading frame, tentatively designated ORF704, which starts from the outside of the Bam7 fragment according to our preliminary DNA sequence analysis of the entire Bam7 fragment (data not shown). The three putative tRNA genes were assigned to $tRNA^{Ser}$ (1629-1716), $tRNA^{Gly}$ (2257-2327), and $tRNA^{Met}$ (2378-2451) judging from their anticodon triplets UGA, GCC, and CAU, respectively; the $tRNA^{Met}$ is most likely an initiator tRNA. The clover-leaf structures of these predicted tRNAs are presented in Fig. 3.

As shown in Fig. 1, the deduced genes are arranged in the order --ORF701

--tRNA^{Ser}(UGA)--ORF702--tRNA^{Gly}(GCC)--tRNA^{Met}(CAU)--ORF703--(ORF704), in which ORF701, ORF702, and tRNA^{Gly}(GCC) have the same polarity (left to right in Fig. 1), while the remainder are in the opposite direction.

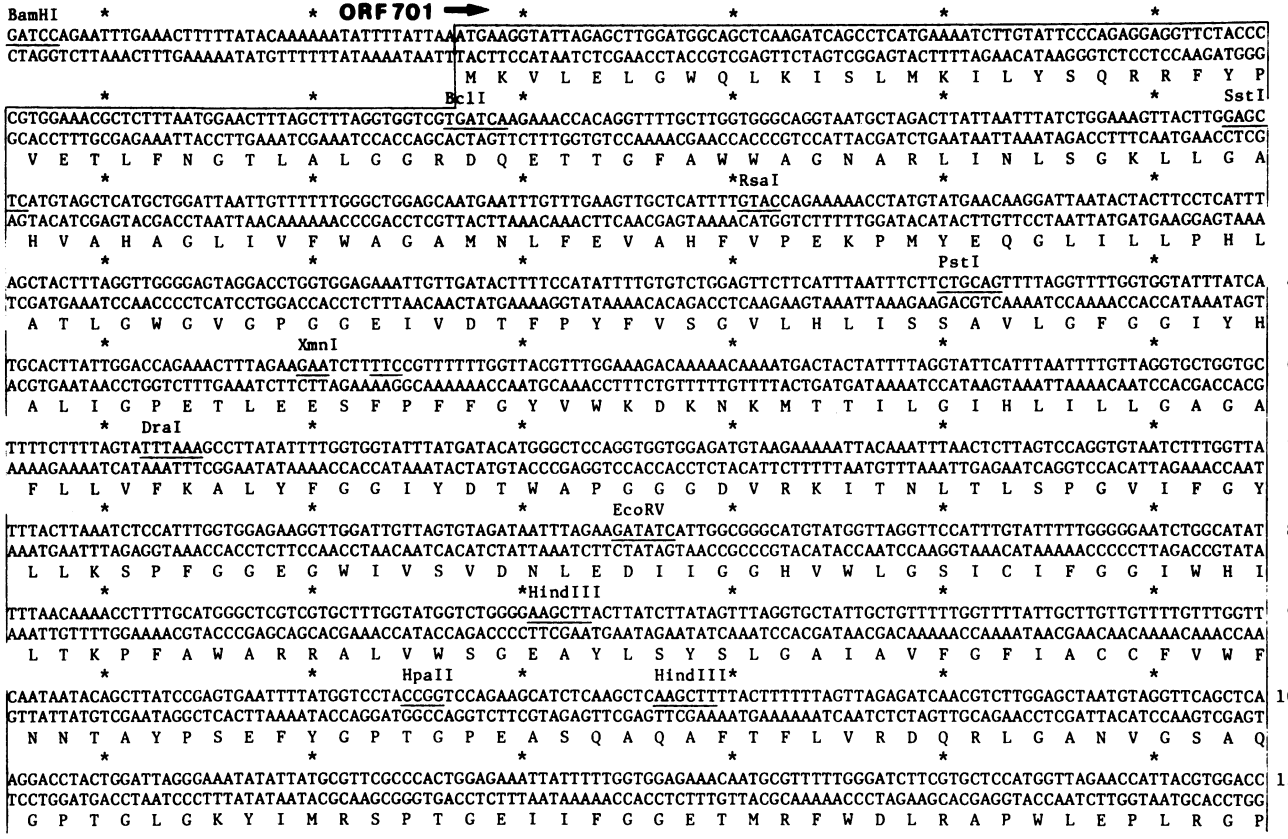
Structural features of the tRNA sequences

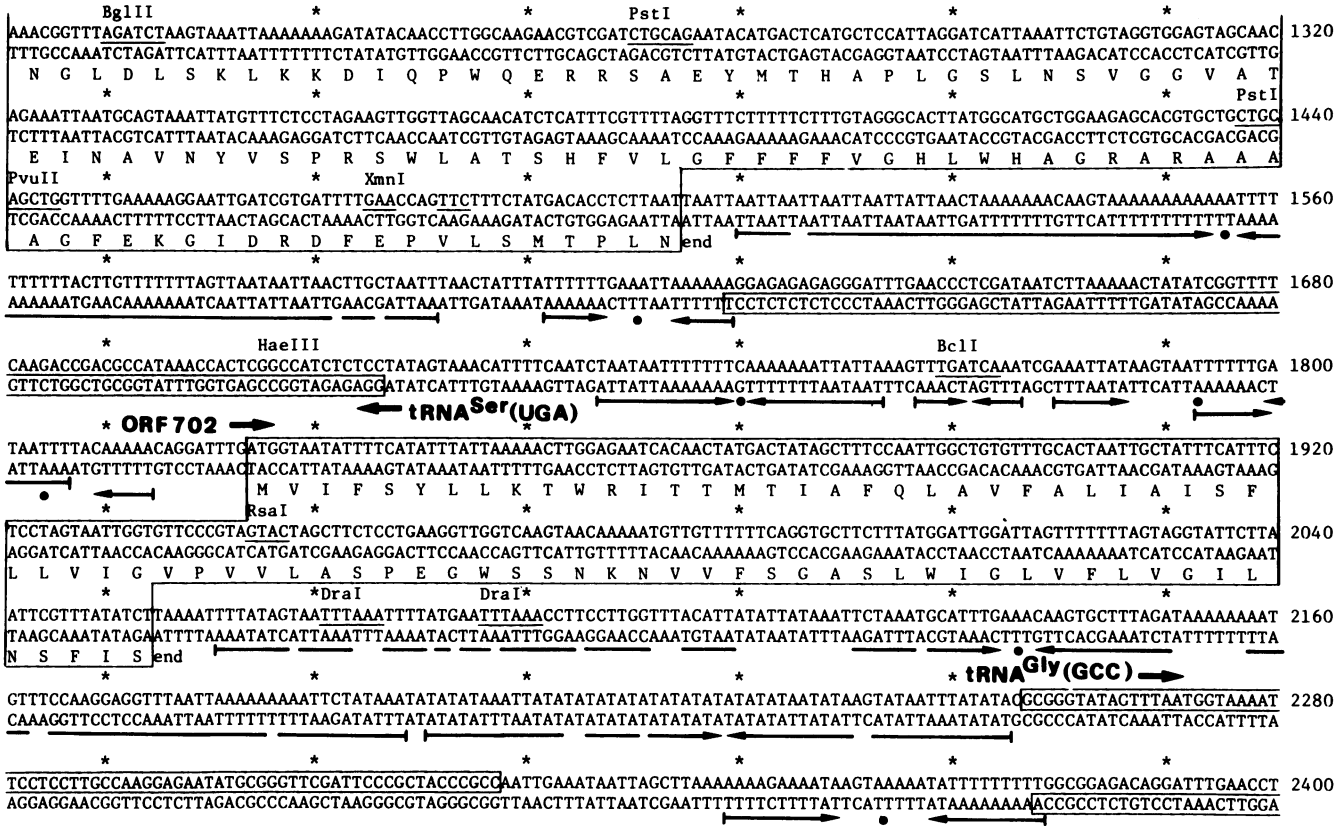
As briefly mentioned above, a stretch of nucleotide sequence between 1629 and 1716 in the lower strands in Fig. 2 can be folded into a clover-leaf structure with the anticodon 5'- UGA -3' which corresponds to serine codons (UCA/G) (Fig. 3a). To our knowledge, this is the first instance of chloroplast tRNA^{Ser} having an anticodon of UGA. However, this putative tRNA^{Ser}(UGA) shows high sequence homologies with maize chloroplast tRNA^{Ser}(GGA) (75%) (9) and tobacco chloroplast tRNA^{Ser}(GCU) (73%) (10). It also shows homologies with Bacillus subtilis tRNA^{Ser}(UGA) (64%) (11) and with E. coli tRNA$^{Ser}_1$(UGA) (60%) (12). Recently, Y. Yamano and K. Ohyama (unpublished results) have detected a corresponding tRNA^{Ser}(UGA) molecule in liverwort chloroplasts, which contained the 3'-CCA terminus and hybridized only with the Bam7 fragment. Thus the DNA stretch above is accounted for as an active tRNA gene in the chloroplasts.

Based on the DNA sequence of 2257-2327 in Fig. 2, a clover-leaf tRNA structure carrying an anticodon GCC, for glycine codons, is presented in Fig. 3b. This putative tRNA^{Gly}(GCC) shows 65%, 81%, 71% and 66% sequence homologies with Euglena gracilis chloroplast tRNA^{Gly}(GCC) (13), tobacco chloroplast tRNA^{Gly}(UCC) (14), B. subtilis tRNA^{Gly}(GCC) (15) and E. coli tRNA^{Gly}(GCC) (12), respectively. In this clover-leaf structure, a mismatched base-pair (C_{26}-A_{42}) is seen, which results in a 4bp anticodon stem instead of the consensus 5bp stem. Such a mispairing, however, has often been encountered in the anticodon stems of chloroplast tRNAs, e.g. C_{26}-A_{42} in tobacco tRNA^{Gly}(UCC) (14), A_{26}-C_{42} in tobacco tRNA^{Gln}(UUG) (9), and A_{27}-A_{43} in tobacco tRNA^{Val}(UAC) (16), or A_{27}-A_{43} in spinach tRNA^{Val}(UAC) (12).

The putative tRNA^{Met}(CAU) gene (positions 2378-2451) may be accounted for as a gene for an initiator rather than an elongator, since the top of the amino-acid acceptor stem in the clover-leaf has the characteristic mismatched base-pair (C_1-C_{73}) of an initiator tRNA^{Met}f (12) (Fig. 3c). Moreover, the sequence of tRNA^{Met}(CAU) is highly homologous to chloroplast and procaryotic initiator tRNAs, e.g. 93% homology with spinach chloroplast (12), 88% with E. gracilis chloroplast (13) and Phaseolus vulgalis chloroplast (12), 85% with Anacystis nidulans (12), and 80% with B. subtilis (12), and E. coli (12).

In the case of E. gracilis chloroplasts, two genes for tRNA^{Gly}(GCC) and initiator tRNA^{Met}(CAU) are very closely linked, with a 12bp spacer, forming a single transcription unit (13). In the present case of liverwort


BamHI
ORF701
BclI
XmnI
DraI
EcoRV
HindIII
HpaII
HindIII
RsaI
PstI
SstI


BglII
PstI
PstI
PvuII
XmnI
HaeIII
BclI
RsaI
DraI
DraI
ORF 702
tRNA Ser (UGA)
tRNA Gly (GCC)

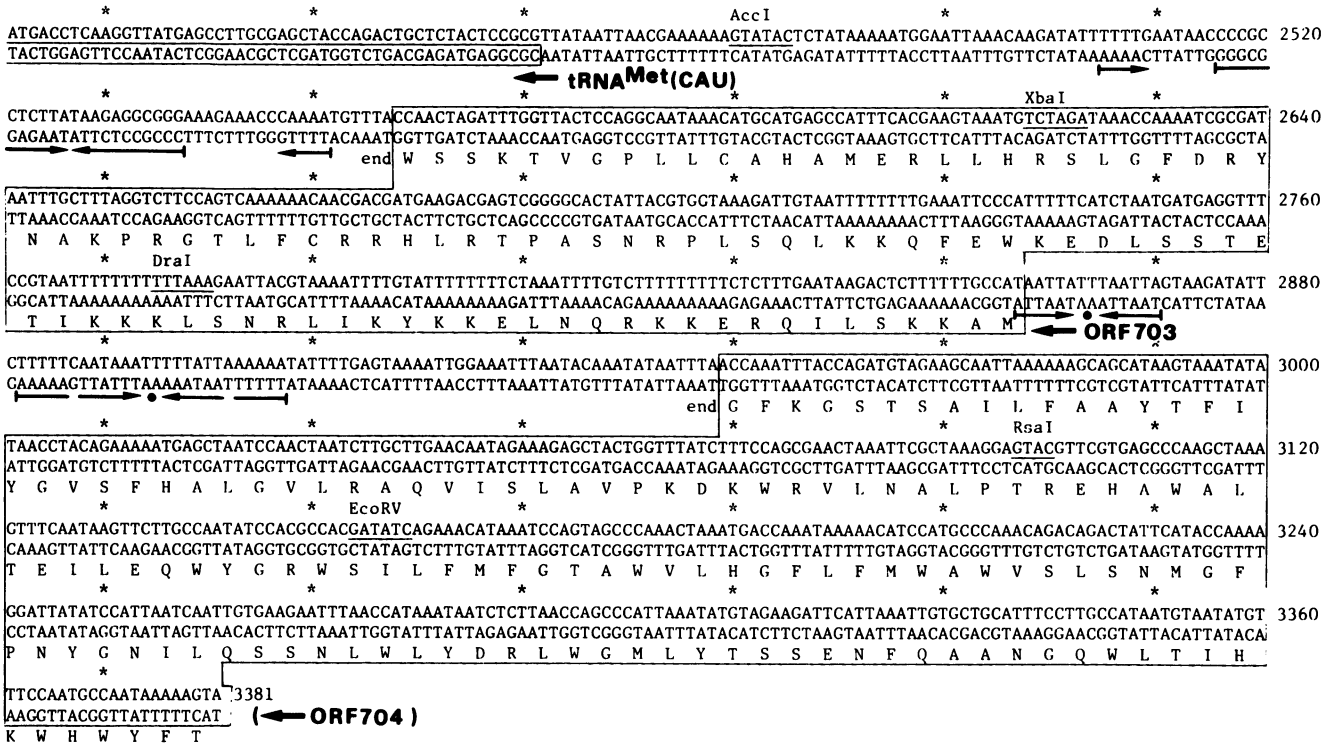


Fig. 2 Nucleotide sequence of the 3.38kbp portion of the Bam7 fragment. Open reading frames and tRNA-like sequences are in solid-lined boxes. The recognition sites of restriction enzymes shown in Figure 1 are underlined. Thick arrows indicate the direction of transcription of the putative genes. Inverted repeats in the spacer regions are indicated by arrows under the sequences.

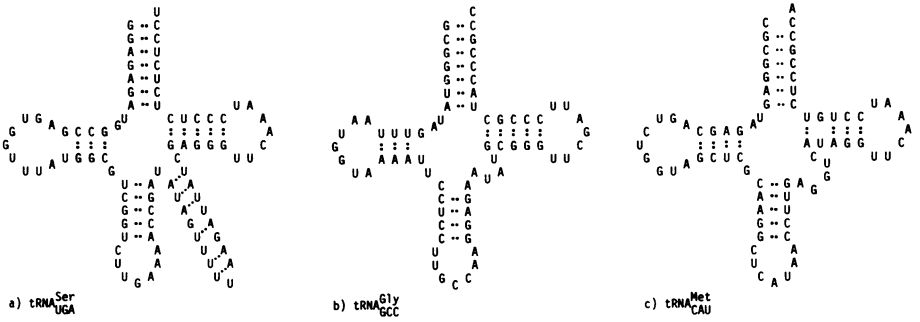


Fig. 3 Clover-leaf structures of the predicted tRNA sequences. Information about base modifications is not available.

chloroplasts, these two tRNA genes are linked with a 50bp spacer, but they are in a tail-to-tail configuration, and therefore, must be transcribed separately in opposite directions, terminating within this spacer region (see below).

All of the three putative tRNA genes detected in the present study contain neither intron nor 3'-CCA sequences. As an interesting feature, the DNA sequences just ahead of these tRNA structural genes are quite similar, namely TTTACTATA-tRNA^{Ser}(UGA), TTTATATAC-tRNA^{Gly}(GCC) and TTAATTATAA-tRNA^{Met}(CAU).

ORF703, a gene encoding a protein homologous to the E. coli ribosomal protein S14

ORF703 contains 99 codons in addition to the initiation codon ATG and terminates at TAA (Fig. 2). The high content of basic amino acid residues (26% with lysine and arginine) of the predicted protein from ORF703 suggests that this protein may play a role in nucleic acid-protein interaction, such as that of a ribosomal protein. We, therefore, searched for amino acid sequence homology between the ORF703 protein and E. coli ribosomal proteins by the method of dot matrix plotting combined with a computer. Among all of the E. coli ribosomal proteins (K. Isono, personal communication), a 30S component S14 (98 amino acid residues) (17) showed a significant homology of nearly 50% to the ORF703 protein. The alignment of amino acid sequences of the ORF703 and E. coli S14 (ES14) proteins is shown in Fig. 4; identical amino acid residues are 45 among 100 residues and conservative replacements are found at 9 positions. Calculated molecular weight of the ORF703 protein is 11,880 daltons, which is comparable to that of ES14 (11,191 daltons) (17). The secondary structures predicted by the method of Chou and Fasman (18) gave essentially the same profiles with the ES14 and ORF703 proteins (data not

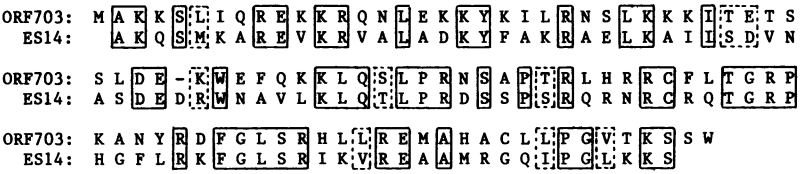


Fig. 4 Comparison of amino acid sequences between ORF703 and E. coli S14 (ES14) proteins. Identical residues are in solid-lined boxes. Broken-lined boxes indicate conservative replacements (D/E, K/R, S/T, L/I/M/V). ES14 sequence was obtained from the analysis of the polypeptide by Yaguchi et al. (17).

shown). Thus, we tentatively conclude that ORF703 represents a gene for a ribosomal protein in liverwort chloroplast which corresponds to S14 in E. coli. Nearly 50% homology in the amino acid sequences has also been reported between E. coli ribosomal proteins and the chloroplast counterparts from various plants; e.g. spinach chloroplast L12 (50% homologous to E. coli L12) (19), tobacco chloroplast S19 (55%) (20), and maize chloroplast S4 (40-50%) (21).

Recently, the complete nucleotide sequence of the E. coli spc operon has been published, which contains ten ribosomal protein genes including the S14 gene (rpsN) (22). The rpsN gene consists of the initiation codon ATG and the following 98 codons which are consistent with the amino acid sequence data (17) except for the C-terminal residue, i.e. Gly from the DNA sequence and Ser from the amino acid sequence. This discrepancy could be due to a difference in the E. coli strains, according to Yaguchi et al. (17). In the case of ORF703, two additional amino acids (Ser and Trp) are observed at the C-terminus (Fig. 4). In this connection, it is interesting to consider that there is a common sequence of 11 nucleotides, 5'-CTAGYTGGTAA-3' (Y = C or T), which exists in the 3' terminal portions of the ORF703 and rpsN genes. If a single base deletion had occurred in the region preceding this sequence in rpsN, such a gene would have a similar C-terminal amino acid sequence to the ORF703 protein, terminating at a newly created TAA stop codon (Fig. 5). In the spc operon of E. coli, all the other nine ribosomal protein genes are known to terminate at UAA (22). The most probable deletion site would be in the stretch of six A residues underlined in Fig. 5 because such sequences are known to be hot spots for frameshift mutations (23).

Structural features of the proteins expected from ORF701 and ORF702

In contrast to ORF703, none of the protein sequences looked at so far shares considerable homology with the other two open reading frames ORF701 and

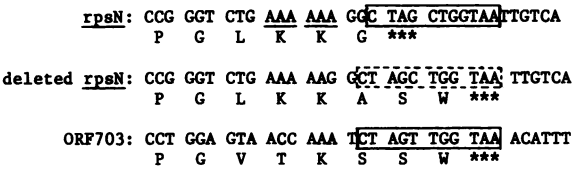


Fig. 5 Homologous DNA sequences (in boxes) at the 3' end regions of rpsN and ORF703. If one of the underlined A residues is deleted from the 3' end sequences of rpsN, the resulting amino acid sequence and stop codon will be "ASW-TAA" as shown in the middle lane (broken box). The DNA sequence of rpsN was obtained from Cerretti et al. (22).

ORF702. The amino acid compositions predicted from them indicate that the proteins have few, if any, cysteines (three in ORF701 and none in ORF702) and high contents of hydrophobic residues (F+L+I+M+V; 33% in ORF701 and 51% in ORF702). The contents of two aromatic residues, Phe and Trp, seem to be high and comparable to those of the psbA protein (24) and P_{680} chlorophyll a apoprotein (25) of spinach, which are related to the photosystem in the plant; 10.9% in ORF701 (F=35, W=18), 13.0% in ORF702 (F=7, W=3), 10.2% in psbA (24) and 11.4% in P_{680} chlorophyll a apoprotein (25). In the case of ORF701, a hydropathy profile by the method of Kyte and Doolittle (26) clearly shows seven hydrophobic clusters and a relatively hydrophilic region as shown in Fig. 6. This profile is highly homologous to that of spinach P_{680} chlorophyll a apoprotein (25). Considering those points, it is tempting to assume that ORF701 represents a membrane protein in chloroplasts (see NOTE ADDED IN PROOF).

Biological effects of ORF701 on E. coli mutant cells

In the course of this study, we noticed an interesting phenomenon, that is, the introduction of pMP227 markedly reduces the degree of streptomycin resistance of the Sm^r (strA) mutants of E. coli, such as HB101, NO1345, or BT52recA. Bacteria with or without the pBR322 plasmid, can grow normally in the presence of Sm (250 μg/ml), while those carrying pMP227 can not grow even at a lower concentration of Sm (100 μg/ml). The plating efficiency of HB101(pMP227) on Sm plates (250 μg/ml) was reduced to the order of 10^{-5}. Further subcloning experiments revealed that, among the ORFs on the Bam7 fragment, ORF701 is responsible for this effect when it is connected to a bacterial promoter at the 5' side. In pMP227, ORF701 had been connected by chance to the Tet^r promoter of pBR322, and this phenomenon was found. A similar effect was also observed when the ORF701 was preceded by the Km^r promoter of pKC7 plasmid (27) in the right orientation. The subcloning

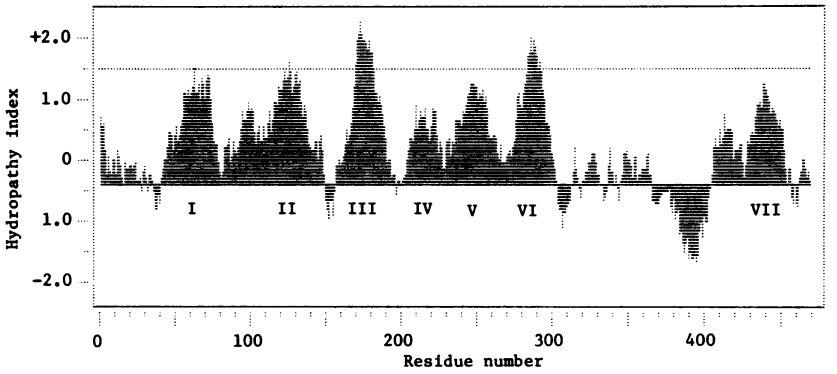


Fig. 6 Hydropathic profile of the ORF701 protein.
According to the method of Kyte and Doolittle (26), the average hydropathic scores of each 19 amino acids are calculated and plotted. Seven clusters of hydrophobic amino acids are designated I to VII.

experiments also revealed that the complete ORF701 sequence is not required, but only a 5'-proximal portion of ORF701 is sufficient to manifest the Sm-sensitive phenotype, since the portion downstream from the HindIII site (positions 891-896, Ala_{284} in Fig. 2) can be removed without any effect. This removable portion represents a relatively hydrophilic region of ORF701 as seen in Fig. 6. The strA (=rpsL, 73 min) is the locus for a 30S ribosomal protein, S12, in E. coli and it is well-known that its Sm^r mutants are recessive to the wild type allele. Accordingly, if a gene corresponding to the bacterial rpsL does exist in the liverwort chloroplast and if it is expressed in E. coli, the introduction of a plasmid carrying such a gene would cause a Sm-sensitive phenotype in the host cells. It is unlikely, however, that ORF701 encodes a counterpart of ribosomal protein S12 (strA) of E. coli, since the deduced amino acid sequence of ORF701 showed no significant homology with that of S12 protein (28). We consider that the ORF701 protein is probably a membrane protein. The role of this protein in E. coli is under investigation.

Codon usage patterns and base composition of the open reading frames

Codon usage patterns of the three open reading frames are quite similar. A strong preference for A or T is observed at the third letters of the synonymous codons. For example, among the six leucine codons, UUA is most frequently used, i.e. 65% in ORF701, 45% in ORF702, and 64% in ORF703; among the four alanine codons, GCC and GCG codons are hardly used (GCC+GCG / four synonyms; ORF701 = 1/42, ORF702 = 0/6, ORF703 = 1/5). A similar tendency is also seen in the portion of presumed ORF704 so far sequenced (see Fig. 2). If

the first and second letters are random for four letters, while all of the third letters are occupied with A or T, then the GC content is expected to be 33.3%. The actual percentages are 37.5% for ORF701, 33.3% for ORF704, 31.6% for ORF702, and 31.0% for ORF703.

The AT-drift is also observed at extra loops of two tRNA genes (AT content in extra loops; 16/19 in $tRNA^{Ser}$(UGA) and 4/4 in $tRNA^{Gly}$(GCC)). GC contents of the tRNA genes are 45% in $tRNA^{Ser}$(UGA), 51% in $tRNA^{Gly}$(GCC) and 54% in $tRNA^{Met}$(CAU), respectively.

Characteristic features of the spacer regions

The spacers separating the ORFs (701-704) and tRNA genes (Ser, Gly, and Met) on the Bam7 fragment are designated I to VII, i.e. -I-(701)-II-(Ser)-III-(702)-IV-(Gly)-V-(Met)-VI-(703)-VII-(704)--. The shorter lengths (50-200bp) of these spacers may be accounted for as a reflection of the relatively compact size of the liverwort chloroplast genome. The most remarkable feature of the spacer regions is its extremely high content of A-T pairs. The average A+T content through spacer I to VII is 85.2% (Fig. 2). Continuous long A+T stretches are often found; in spacer IV, a nearly 50 A+T stretch including a $(TA)_{12}$ sequence is observed (Fig. 2).

In addition to the abundance of AT pairs, the spacer regions contain many inverted repeats in their nucleotide sequences (Fig. 2). In spacer II, the sequence from 1510 to 1601 can be presented in a large stem-and-loop configuration with a perfect 35bp inverted repeat that splits the 3' ends of the coding sequences of both ORF701 and $tRNA^{Ser}$(UGA) genes (Fig. 7a). Spacer III contains a number of small inverted repeats at positions 1737-1763, 1767-1776, 1780-1806, and 1794-1814 (Fig. 2). The longest spacer IV may include two large stem-and-loop structures; one is located at positions 2061-2198, immediately following the ORF702 stop codon, and the other at positions 2201-2255, just ahead of the $tRNA^{Gly}$(GCC) gene (Fig. 7b). In spacer V, both the upper and lower strands of positions 2349-2379 can be presented in similar stem-and-loop structures that are followed by a short T-stretch (Fig. 7c). Thus this region may serve as a dual terminator for both the $tRNA^{Gly}$(GCC) gene and the $tRNA^{Met}$(CAU) gene, which are transcribed in opposite directions (Fig. 1). In spacer VI, an inverted repeat at positions 2505-2551 close to the 3' end of ORF703 is somewhat different from those in the other spacers, as it is composed of a GC-rich stem (Fig. 2). Spacer VII also contains two inverted repeats at positions 2857-2870 and 2882-2908 (Fig. 2).

Genetic signals for transcription, translation, and processing, such as promoters, terminators, and others are not assigned directly in this study,

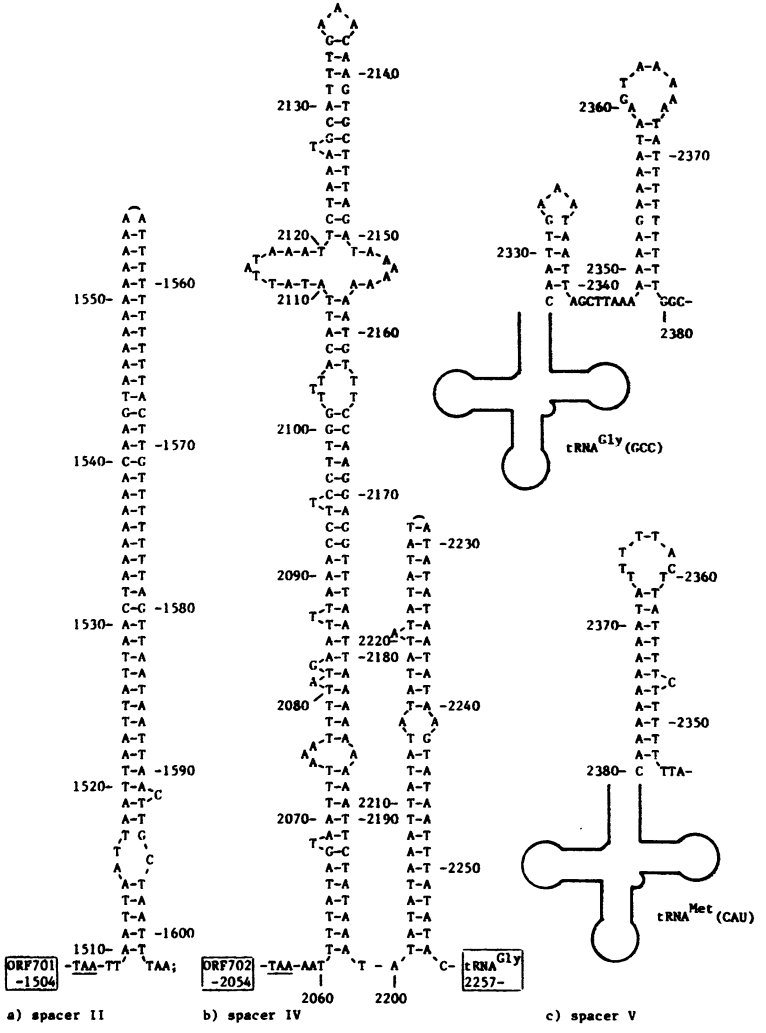


Fig. 7 Possible stem-and-loop structures in the spacer regions.

but they must be located within the relatively short spacers of AT-rich sequences with many inverted repeats. Although we analyzed only a portion of the central part of the liverwort chloroplast genome, the punctuation of the genes by those characteristic spacers might be a general feature of the liverwort chloroplast genome.

AKNOWLEDGEMENTS

We are grateful to Dr. Yoshiro Shimura for his critical reading of the manuscript. This work was supported by a Grant-in-Aid for Special Project Research from the Ministry of Education, Science, and Culture of Japan.

*To whom correspondence should be addressed

REFERENCES

1. Ohyama, K., Yamano, Y., Fukuzawa, H., Komano, T., Yamagishi, H., Fujimoto, S. and Sugiura, M. (1983) Mol. Gen. Genet. 189, 1-9.
2. Whitfeld, P. R. and Bottomly, W. (1983) Ann. Rev. Plant Physiol. 34, 279-310.
3. Ohyama, K., Wetter, L. R., Yamano, Y., Fukuzawa, H. and Komano, T. (1982) Agric. Biol. Chem. 46, 237-242.
4. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory
5. Jaskunas, S. R., Fallon, A. M. and Nomura, M. (1977) J. Biol. Chem. 252, 7323-7336.
6. Oka, A., Nomura, N., Morita, M., Sugisaki, K., Sugimoto, K. and Takanami, M. (1979) Mol. Gen. Genet. 172, 151-159.
7. Inokuchi, H., Yamao, F., Sakano, H. and Ozeki, H. (1979) J. Mol. Biol. 132, 649-662.
8. Maxam, A. M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
9. Steinmetz, A. A., Krebbers, E. T., Zsuzsanna, S., Gubbins, E. J. and Bogorad, L. (1983) J. Biol. Chem. 258, 5503-5511.
10. Deno, H. and Sugiura, M. (1983) Nucleic Acids Res. 11, 8407-8414.
11. Green C. J. and Vold, B. S. (1983) Nucleic Acids Res. 11, 5763-5774.
12. Gauss, D. H. and Sprinzl, M. (1983) Nucleic Acids Res. 11, r1-r53.
13. Karabin, G. D. and Hallick, R. B. (1983) J. Biol. Chem. 258, 5512-5518.
14. Deno, H. and Sugiura, M. (1984) Proc. Natl. Acad. Sci. USA 81, 405-408.
15. Wawrousek, E. F. and Hansen, J. N. (1983) J. Biol. Chem. 258, 291-298.
16. Deno, H., Kato, A., Shinozaki, K. and Sugiura, M. (1982) Nucleic Acids Res. 10, 7511-7520.
17. Yaguchi, M., Roy, C., Reithmeier, R. A. F., Wittmann-Liebold, B. and Wittmann, H. G. (1983) FEBS Lett. 154, 21-30.
18. Chou, P. Y. and Fasman, G. D. (1978) Adv. Enzymol. 47, 45-148.
19. Bartsch, M., Kimura, M. and Subramanian, A.-R. (1982) Proc. Natl. Acad. Sci. USA 79, 6871-6875.
20. Sugita, M. and Sugiura, M. (1983) Nucleic Acids Res. 11, 1913-1918.
21. Subramanian, A.-R., Steinmetz, A. and Bogorad, L. (1983) Nucleic Acids Res. 11, 5277-5286.
22. Cerretti, D. P., Dean, D., Davis, G. R., Bedwell, D. M. and Nomura, M. (1983) Nucleic Acids Res. 11, 2599-2616.
23. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.
24. Zurawski, G., Bohnert, H. J., Whitfeld, P. R. and Bottomley, W. (1982) Proc. Natl. Acad. Sci. USA 79, 7699-7703.
25. Morris, J. and Herrmann, R. G. (1984) Nucleic Acids Res. 12, 2837-2850.
26. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
27. Nagaraja Rao, R. and Rogers, S. G. (1979) Gene 7, 79-82.
28. Funatsu, G., Yaguchi, M. and Wittmann-Liebold, B. (1977) FEBS Lett. 73, 12-17.

NOTE ADDED IN PROOF

Recently, ORF701-like DNA sequence has been reported in the upstream region (1-1429 according to their nucleotide numberings) of psbA locus of E. gracilis chloroplast DNA by Keller & Stutz (FEBS Lett. 175, 173-177, 1984). They discussed two possibilities for this region; (1) two independent open reading frames or (2) the 3'-terminal part of a split gene. Our sequence data support the latter because the spliced message (the coding region will be 17-298, 602-653, and 1077-1429) bears an information for a polypeptide which shares 81.5% homology with the C-terminal portion of ORF701 analyzed in this paper.