Coding sequences for chloroplast ribosomal protein S12 from the liverwort, *Marchantia polymorpha*, are separated far apart on the different DNA strands

Hideya Fukuzawa, Takayuki Kohchi, Hiromasa Shirai, Kanji Ohyama*, Kazuhiko Umesono+, Hachiro Inokuchi+ and Haruo Ozeki+

Laboratory of Plant Molecular Biology, Research Center for Cell and Tissue Culture, Faculty of Agriculture and +Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Received 21 January 1986

During the nucleotide sequencing of chloroplast DNA from the liverwort, *M. polymorpha*, we found that a coding sequence corresponding to the *Esherichia coli* ribosomal protein S12 gene (*rps*12) is split into three exons. Strikingly, the first exon with the 5'-intron boundary sequence was found on the opposite strand of the chloroplast DNA (120 kilobases long, circular molecules) approx. 60 kilobases away from the rest of the exons. The amino acid sequence deduced from the DNA sequence was highly homologous to the sequences of the S12 ribosomal protein of *E. coli* (70.2%), and *Euglena gracilis* chloroplasts (73.6%). Possible mechanisms for the expression of this split gene are discussed.

Chloroplast *Ribosomal protein S12* *Intron* *Trans-splicing* (Marchantia polymorpha)

1. INTRODUCTION

Introns (intervening sequences) in a chloroplast RNA gene have been reported: the 23 S rRNA gene of *Chlamydomonas reinhardii* [1]; the tRNA genes, *trn*I(GAU) and *trn*A(UGC), in the 16 S–23 S rDNA spacer region of *Zea mays* [2] and *Nicotiana tabacum* [3]; as well as the chloroplast tRNA genes *trn*L(UAA) [4,5], *trn*K(UUU) [6], *trn*G(UCC) [7,8] and *trn*V(UAC) [9–11]. Introns within a chloroplast protein gene have also been reported in several genes of *Euglena gracilis*; for the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbc*L) [12], the elongation factor Tu (*tuf*A) [13], and the 32 kDa protein (*psb*A) [14,15]. The gene for the 32-kDa protein of *C. reinhardii* also has introns [16] as does the gene for the H^+-ATPase subunit I (*atp*F) of wheat [17]. Zurawski et al. [18] reported that the chloroplast ribosomal protein L2 (*rpl*2) in *N. debneyi* has a single intron. We have also detected several genes with introns in the chloroplast DNA from the liverwort, *Marchantia polymorpha* (unpublished).

Recently, Hallick et al. [19] reported that the reading frame of the ribosomal protein S12 in *N. tabacum* is interrupted by two introns, but described only the second one. During nucleotide sequencing of chloroplast DNA from the liverwort, *M. polymorpha*, however, we found the first exon with the 5'-intron boundary sequence on the opposite strand of the chloroplast DNA. Here, we present the complex structure of the putative gene for chloroplast ribosomal protein S12 from *M. polymorpha* which has 3 exons split into different DNA strands.

2. MATERIALS AND METHODS

Chloroplasts were prepared from cell suspension cultures of *M. polymorpha* as in [20]. Chloroplast DNA fragments, the *Bam*HI (Ba11) and *Bgl*II

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

(Bg5) fragments, were cloned into the respective plasmids, pBR322 and pKC7, as described in [21]. A physical map of the chloroplast DNA for *Bam*HI fragments has been described in [22,23]. The location of the Bg5 fragment on that map was determined by restriction analysis and Southern hybridization (fig.1). For DNA sequencing, the recombinant plasmids were sonicated and cloned randomly into the *Hinc*II site of the M13mp18 vector as described in [24]. Recombinant phages containing chloroplast DNA fragments were selected by filter hybridization with ^{32}P-labeled chloroplast DNA [25]. DNA sequencing was done by the dideoxy method [26] then analyzed by a computer (Hitachi DNASIS system).

3. RESULTS AND DISCUSSION

We first identified a coding region for ribosomal protein S12 on the *Bam*HI fragment (Ba11) using Southern hybridization with an *Eu. gracilis* probe

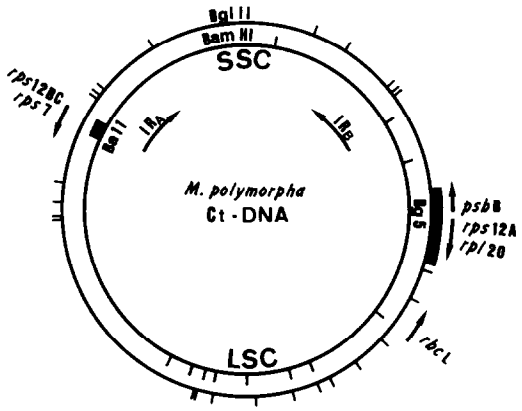

Fig.1. Locations of coding regions for chloroplast ribosomal protein S12 on maps of chloroplast DNA from the liverwort, *M. polymorpha*. Exon 1 (*rps*12A) was located on the *Bgl*II fragment (Bg5); the direction of its transcription was clockwise (arrow). By contrast, exons 2 and 3 (*rps*12B and C) were found on the *Bam*HI fragment (Ba11), their transcription being in the opposite direction from that of exon 1 (arrow). *rps*12, *rps*7 and *rpl*20: respective genes of ribosomal proteins S12, S7 and L20. The site of the large subunit gene (*rbc*L) for ribulose-1,5-bisphosphate carboxylase is shown. IR_A and IR_B indicate a set of inverted repeats, and SSC and LSC the small single copy and large single copy regions.

(provided by Drs Montandon and Stutz [27]). The Ba11 fragment was mapped at the junction between the inverted repeat (IR_A) and the large single copy (LSC) region (fig.1). DNA sequence analysis revealed, however, that the sequence for the N-terminal 38 amino acids of the protein was missing in this coding region.

We searched our nucleotide sequence data files for the missing N-terminal 38-amino-acid sequence and found it on the *Bgl*II fragment (Bg5) approx. 60 kilobases away on the opposite DNA strand (fig.1). Complete nucleotide sequences of the coding regions for ribosomal protein S12, *rps*12A and *rps*12B-C, including the flanking regions, are shown in fig.2. Exons 1 and 2 were followed by a consensus sequence (GTGCG) of the 5′-boundary regions of the introns found in chloroplast genes [6,19]. We also found a much conserved consensus sequence, AAGCCG..TGAA...AAA...TCA.G-T.CGGTTT, in the introns 75 nucleotides upstream from exon 2 and 61 nucleotides upstream from exon 3. This consensus sequence has been present in all the introns found so far in chloroplast genes of *Eu. gracilis* [14]. A 'connecting' helix (underlined with arrows, fig.2B) is formed 2–3 nucleotides upstream from the exons in the 3′-intron boundary regions [14].

The gene organizations deduced from DNA sequences near coding regions for ribosomal protein S12 are shown in fig.3. Open reading frames corresponding to the ribosomal proteins S7 and L20 were identified by their amino acid sequence homologies, 42.6 and 44.4%, for the respective *E. coli* ribosomal proteins [28,29]. We reported earlier that the chloroplast ribosomal protein S14 from *M. polymorpha* has 45.0% homology to that of *E. coli* [23]. By contrast, the amino acid sequence of chloroplast ribosomal protein S12 from *M. polymorpha* showed markedly higher homologies, 73.6% to that of *Eu. gracilis* [27] and 70.2% to that of *E. coli* [29] (fig.4). The amino acid sequence of ribosomal protein S12 from *M. polymorpha* near the splicing junctions (arrowheads, fig.4) showed an even higher homology to sequences from *Eu. gracilis* and *E. coli*, both of which have no intron [27,29]. This highly conserved amino acid sequence suggests that the chloroplast ribosomal protein S12 may play an essential part in the ribosomal function during protein synthesis in chloroplasts.

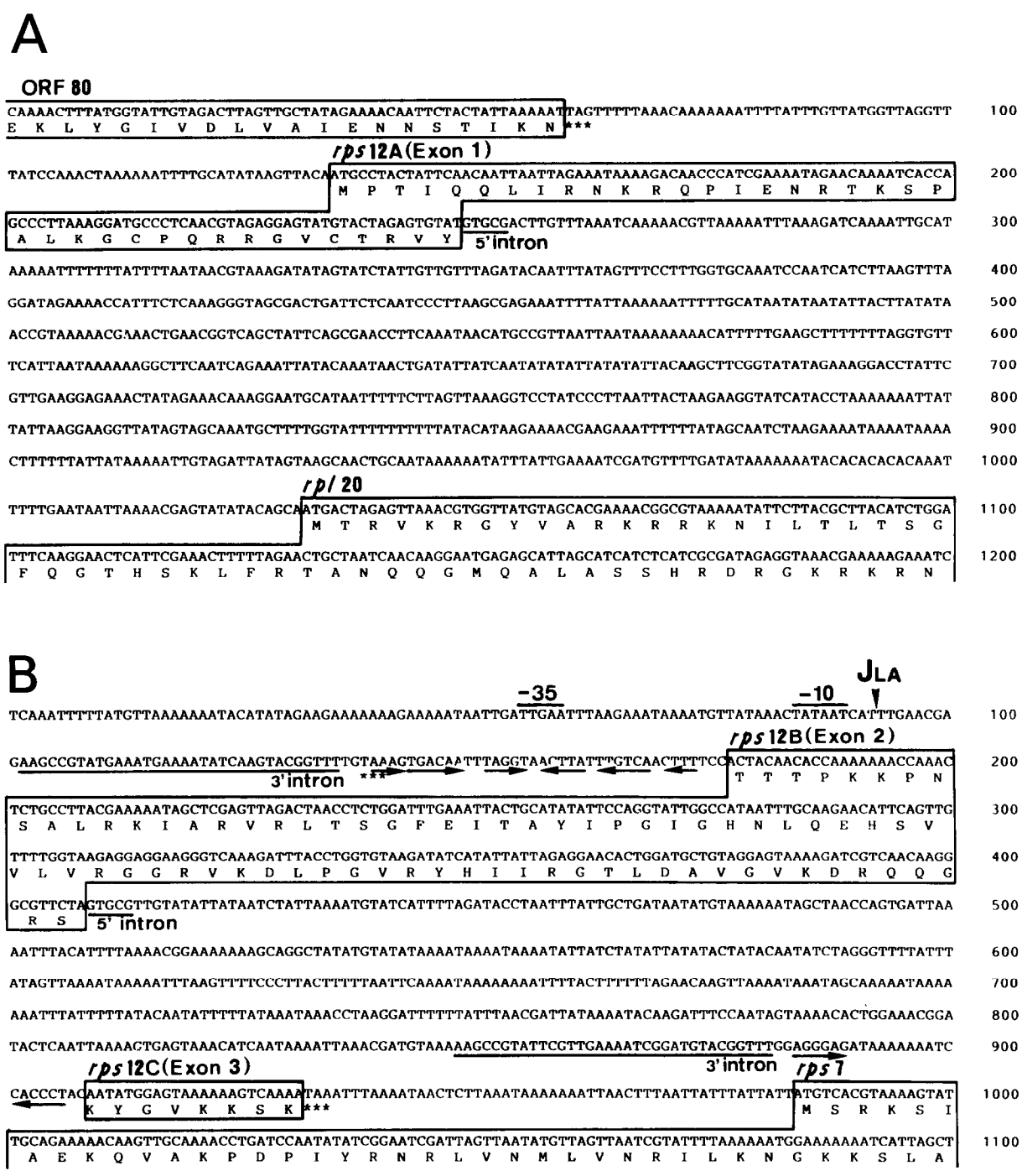

Fig.2. Complete nucleotide sequences near exon 1 (A), and exons 2 and 3 (B) of the *rps*12 gene. The exons are boxed, and the consensus sequences of their 5′-intron boundary regions, as well as those of the near 3′-intron boundary regions, underlined. Connecting helices are shown by underlining arrows. J_{LA} denotes the junction of an inverted repeat (IR_A) and a large single copy region (LSC). Amino acids are expressed using the one-letter code.

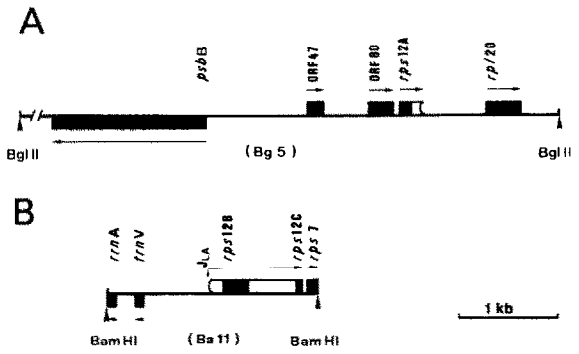

Fig.3. Gene organizations near the *rps*12 gene in exon 1 (A) and exons 2 and 3 (B). Arrows indicate the direction of transcription. *rps*12, *rps*7 and *rpl*20 as in fig.1. *psb*B, gene for the P680 protein in photosystem II. J_{LA}, junction of an inverted repeat (IR_A) and a large single copy region (LSC).

Two possible open reading frames (ORF80 and ORF47) were detected further upstream from exon 1 (fig.3A). Exon 1 of ribosomal protein S12 with the 5'-intron boundary sequence was followed by a coding sequence of ribosomal protein L20 (figs 2A,3A). Following the coding region of ribosomal protein S12 in exon 3 is a ribosomal protein S7 coding region in exon 3 (figs 2B,3B). Close linkage of ribosomal protein S7 and S12 genes also exists in *Eu. gracilis* [27] and *E. coli* [29].

Transcription for exons 2 and 3, as well as for the ribosomal protein S7 gene, is initiated by a typical prokaryotic promoter sequence (−35 and −10 regions) found upstream (fig.2B). S_1 mappings showed that this promoter was highly active in chloroplasts as well as in *E. coli* [30]. Northern hybridizations also showed the active transcription for exon 1 (not shown). If the split gene described here provides active mRNA, there must be a rejoining of exons at the RNA or DNA level. Results of S_1 mappings and Northern hybridizations suggest transcription units for exons 2 and 3 that are independent of the unit for exon 1. Therefore, active mRNA for ribosomal protein S12 probably is formed post-transcriptionally by a mechanism such as that of trans-splicing described in [31,32]. An investigation of the transcription and splicing mechanisms for the split gene described here is now in progress.

ACKNOWLEDGEMENT

This research was supported in part by a Grant-in-Aid for Special Research Projects from the Ministry of Education, Science, and Culture, Japan. We thank Professor Yasuyuki Yamada, Director of the Research Center for Cell and Tissue Culture, Faculty of Agriculture, Kyoto University, for his continuous encouragement during this study. We also thank Drs P.E. Montandon and E. Stutz for providing us with the *Eu. gracilis* probe.

```
                    10        20        30        40        50        60
                    +         +         +         +         +         +
E. gracilis         MPTLEHLTRSPRKKIKRKTKSPALKGCPQKRAICMRVYTTTPKKPNSALRKVTRVRLSSG
                    ***   * *  *  *   *********** *  * ********************  **** **
M. polymorpha       MPTIQQLIRNKRQPIENRTKSPALKGCPQRRGVCTRVYTTTPKKPNSALRKIARVRLTSG
                    * *  ** *  *         ***  *** **********************  ***** *
E. coli             MATVNQLVRKPRARKVAKSNVPALEACPQKRGVCTRVYTTTPKKPNSALRKVCRVRLTNG

                    70        80        90        100       110       120
                    +         +         +         +         +         +
E. gracilis         LEVTAYIPGIGHNLQEHSVVLIRGGRVKDLPGVKYHVIRGCLDAASVKNRKNARSKYGVKKPKPK
                     * ****************** ********** ** *** ***  ** *  ********* *
M. polymorpha       FEITAYIPGIGHNLQEHSVVLVRGGRVKDLPGVRYHIIRGTLDAVGVKDRQQGRSKYGVKKSK
                    ** * ** * ********* * **************  ** **  ***** * *******  *
E. coli             FEVTSYIGGEGHNLQEHSVILIRGGRVKDLPGVRYHYVRGALDCSGVKDRKQARSKYGVKRPKA
```

Fig.4. Amino acid sequences for the ribosomal protein S12 from *M. polymorpha* compared with those from *E. coli* and *Eu. gracilis*. Vertical arrowheads indicate sites of splicing junctions in the ribosomal protein S12 from *M. polymorpha*. Asterisks denote amino acids that are identical between the two proteins. Amino acids are expressed in the one-letter code.

REFERENCES

[1] Rochaix, J.D. and Malnoe, P. (1978) Cell 15, 661–670.

[2] Koch, W., Edwards, K. and Kossel, H. (1981) Cell 25, 203–213.

[3] Takaiwa, F. and Sugiura, M. (1982) Nucleic Acids Res. 10, 2665–2676.

[4] Steinmetz, A., Gubbins, E.J. and Bogorad, L. (1982) Nucleic Acids Res. 10, 3027–3037.

[5] Bonnard, G., Michel, F., Weil, J.H. and Steinmetz, A. (1984) Mol. Gen. Genet. 194, 330–336.

[6] Sugita, M., Shinozaki, K. and Sugiura, M. (1985) Proc. Natl. Acad. Sci. USA 82, 3557–3561.

[7] Deno, H. and Sugiura, M. (1984) Proc. Natl. Acad. Sci. USA 81, 405–408.

[8] Quigley, F. and Weil, J.H. (1985) Curr. Genet. 9, 495–503.

[9] Deno, H., Kato, A., Shinozaki, K. and Sugiura, M. (1982) Nucleic Acids Res. 10, 7511–7520.

[10] Krebbers, E., Steinmetz, A. and Bogorad, L. (1984) Plant Mol. Biol. 3, 13–20.

[11] Zurawski, G. and Clegg, M.T. (1984) Nucleic Acids Res. 12, 2549–2559.

[12] Koller, B., Gingrich, J.C., Stiegler, G.L., Farley, M.A., Delius, H. and Hallick, R.B. (1984) Cell 36, 545–553.

[13] Montandon, P.E. and Stutz, E. (1983) Nucleic Acids Res. 11, 5877–5892.

[14] Keller, M. and Michel, F. (1985) FEBS Lett. 179, 69–73.

[15] Karabin, G.D., Farley, M. and Hallick, R.B. (1984) Nucleic Acids Res. 12, 5801–5812.

[16] Erickson, J.M., Rahire, M. and Rochaix, J.D. (1984) EMBO J. 3, 2753–2762.

[17] Bird, C.R., Koller, B., Auffret, A.D., Huttly, A.K., Howe, C.J., Dyer, T.A. and Gray, J.C. (1985) EMBO J. 4, 1381–1388.

[18] Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1984) Nucleic Acids Res. 12, 6547–6558.

[19] Hallick, R.B., Gingrich, J.C., Johanningmeier, U. and Passavant, C.W. (1985) in: Molecular Form and Function of the Plant Genome (Vloten-Doting, L. et al. eds) pp.211–220, Plenum, New York.

[20] Ohyama, K., Wetter, L.R., Yamano, Y., Fukuzawa, H. and Komano, T. (1982) Agric. Biol. Chem. 46, 237–242.

[21] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning, pp.1–16, Cold Spring Harbor Laboratory, NY.

[22] Ohyama, K., Yamano, Y., Fukuzawa, H., Komano, T., Yamagishi, H., Fujimoto, S. and Sugiura, M. (1983) Mol. Gen. Genet. 189, 1–9.

[23] Umesono, K., Inokuchi, H., Ohyama, K. and Ozeki, H. (1984) Nucleic Acids Res. 12, 9551–9565.

[24] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309–321.

[25] Hu, N.-T. and Messing, J. (1982) Gene 17, 271–277.

[26] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.

[27] Montandon, P.E. and Stutz, E. (1984) Nucleic Acids Res. 12, 2851–2859.

[28] Wittmann-Liebold, B. and Seib, C. (1979) FEBS Lett. 103, 61–65.

[29] Post, L.E. and Nomura, M. (1980) J. Biol. Chem. 255, 4660–4666.

[30] Fukuzawa, H., Uchida, Y., Yamano, Y., Ohyama, K. and Komano, T. (1985) Agric. Biol. Chem. 49, 2725–2731.

[31] Solnick, D. (1985) Cell 42, 157–164.

[32] Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) Cell 42, 165–171.