Volume 16 Number 21 1988 Nucleic Acids Research

A nicked group II intron and *trans*-splicing in liverwort, *Marchantia polymorpha*, chloroplasts

Takayuki Kohchi[1,2], Kazuhiko Umesono[3], Yutaka Ogura[1], Yuriko Komine[3], Kenji Nakahigashi[3], Tohru Komano[2], Yasuyuki Yamada[1], Haruo Ozeki[3] and Kanji Ohyama[1]*

[1]Research Center for Cell and Tissue Culture, [2]Department of Agricultural Chemistry, Faculty of Agriculture and [3]Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Received August 16, 1988; Accepted September 27, 1988

ABSTRACT

The chloroplast gene rps12 for ribosomal protein S12 in a liverwort, Marchantia polymorpha, is split into three exons by two introns, one of which (intron 1) is discontinuous. Exon 1 of rps12 for the N-terminal portion of the S12 protein is far from exons 2 and 3 for the C-terminal portion on the opposite DNA strand. S1-nuclease protection analysis and Northern hybridization with RNA isolated from the liverwort chloroplasts showed that: (i) the exons 1 and 2-3 of the rps12 gene with the neighboring genes were transcribed separately, (ii) the trans-splicing of intron 1 occurred after the processing of two primary transcripts to two pre-mRNAs, and (iii) there was no particular order for the splicing of intron 1 (trans) and intron 2 (cis) in the rps12 gene. We propose a bimolecular interaction model for trans-splicing by assuming that intermolecular base pairings between two pre-mRNAs result in the formation of the structure typical of group II introns except for disruption in the loop III region. This structure could be constructed in intron 1 of tobacco rps12 gene.

INTRODUCTION

A number of chloroplast genes are interrupted by introns and thus require post-transcriptional RNA splicing for the gene expression (1-4). Chloroplast introns belong to either group I or group II depending on their secondary structures, which was first elucidated in mitochondria (5). Both of these intron families are accounted as self-splicing types, which suggests that catalytic activity resides in the intron RNA itself even if some protein factors are required for efficient splicing. We have detected 20 different introns in the chloroplast genome of a liverwort, Marchantia polymorpha; only one intron, in the trnL(UAA) gene, belongs to group I, and the other 19 in group II (1,2). Among these split genes, we found an unusual organization for the rps12 gene that encodes the 30S ribosomal protein S12; the gene consists of three exons and two introns, and exon 1 is far (some 60 kb) from the other two exons on the opposite DNA strand (6). Figure 1 illustrates the genes neighboring the two separated parts, designated as rps12' (A) and rps'12 (B). Essentially the same gene organization as that of rps12 has also been reported for the tobacco

chloroplast genome, although the portion of gene containing exons 2-3 (rps'12) is duplicated in the inverted-repeat regions (4, 7-10). This gene organization suggests that the two parts of rps12 are transcribed separately and then spliced trans to assemble exons 1 and 2. Electron microscopic analysis of RNA-DNA hybrids in tobacco showed that there are separate transcripts of exon 1 and 2-3 of the rps12 gene in the chloroplast as well as spliced RNA molecules in which exon 1 is joined to exon 2 (11). The gene psaA1 for photosystem I P700 protein in Chlamydomonas reinhardtii chloroplasts has been reported to be a discontinuous gene split into three separated locations, suggesting the operation of trans-splicing (12, 13).

Two separate RNA molecules can be joined by trans-splicing in vitro in nuclear extracts of HeLa cells (14, 15). In yeast, efficient trans-splicing in vitro of the mitochondrial group II intron suggests that interaction between the 3'-end of the 5'-exon and intron is needed (16). Recently several cases of trans-splicing in vivo have been reported (see review: 17). Unlike these cases, however, the chloroplast gene for ribosomal protein S12 involved the trans-splicing of the coding region for a single protein.

In this paper, we identified various RNA molecules derived from the two regions of the liverwort chloroplast genome shown in Figure 1. We postulated maturation pathways for mRNAs from the primary RNA transcripts via RNA processing and splicing cis and trans, and propose a bimolecular interaction model for trans-splicing according to the folding model of group II introns of Michel and Dujon (5).

MATERIALS AND METHODS

Preparation of liverwort chloroplast RNA: Liverwort chloroplasts were isolated from the cells of two-week-old suspension cultures grown under continuous illumination (18). Chloroplast RNA was prepared as described before (19).

Northern hybridization analysis: Oligodeoxyribonucleotide probes (16 nucleotides long) were synthesized with a DNA synthesis apparatus (Shimadzu NS-1) and labeled with [γ-^{32}P]ATP (Amersham, 5000 Ci/mmol) and T4 polynucleotide kinase (Takara Shuzo). RNA probes were prepared with in vitro transcription of SP6 vectors containing liverwort chloroplast DNA fragments by the use of [α-^{32}P]UTP (Amersham, 800 Ci/mmol). Northern hybridization was done at 45°C for oligonucleotide probes and in the presence of 50% formamide for riboprobes by a procedure described previously (19).

S1-nuclease protection analysis: DNA probes for S1-nuclease protection analysis were obtained from recombinant plasmids containing liverwort chloroplast DNA.

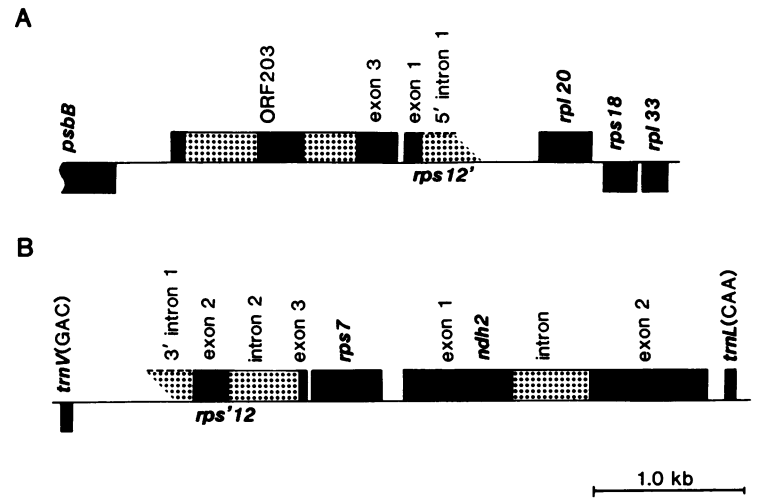


Figure 1. Gene organization of the rps12 gene. (A) Gene organization in the region coding for rps12'. Symbols are: psbB for the 47-kDa chlorophyll a apoprotein in photosystem II; ORF203 for an open reading frame of 203 amino acids; and rps12, rpl20, rps18, and rpl33 for ribosomal proteins S12, L20, S18, and L33, respectively. (B) Gene organization of the region coding for rps'12. Symbols: rps7 for ribosomal protein S7; ndh2 for NADH dehydrogenase (ND2); and trnV(GAC) and trnL(CAA) for valine tRNA(GAC) and leucine tRNA(CAA), respectively. Genes shown above the lines are transcribed to the right, and those under the lines are transcribed to the left. Hatched boxes are introns.

The probes were labeled at the 5'-end as described above and at the 3'-end with [α-^{32}P]dCTP (Amersham, 3000 Ci/mmol) and Klenow fragment of DNA polymerase I (Takara Shuzo). S1-nuclease protection analysis was done by a published procedure (20).

RESULTS

Primary RNA transcripts of rps12' and rps'12 and the processing products.

The 5'-portion of rps12 gene (rps12') was proceeded by an unidentified ORF203 that carried two introns, and followed by the rpl20 gene for the 50S ribosomal protein L20 (Fig. 1A). The remaining 3'-portion of rps12 (rps'12) was followed by the rps7 gene for the 30S ribosomal protein S7, ndh2 (a counterpart of human mitochondrial ND2), and trnL(CAA) for the leucine tRNA(CAA) in this order (Fig. 1B). Chloroplast RNAs were extracted from the cultured liverwort cells liverwort, and Northern hybridization experiments were done with appropriate probes. Two separate transcripts for rps12' and rps'12 were

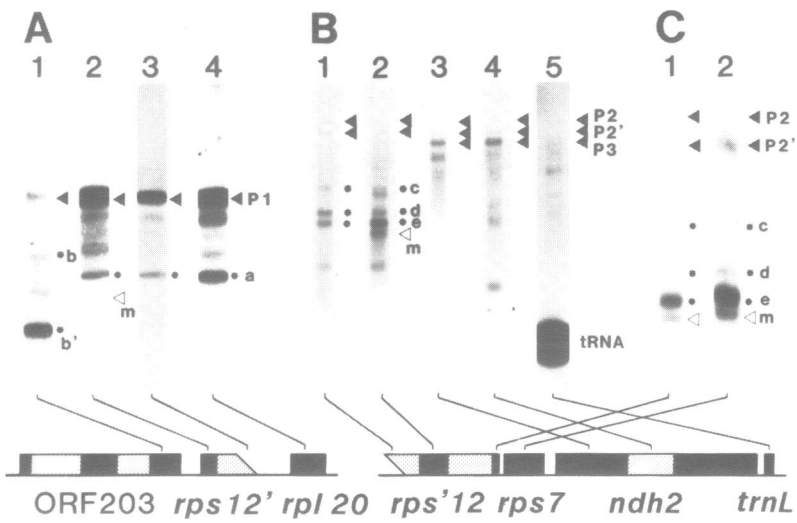


Figure 2. Northern hybridization with a variety of probes. (A) Probes specific to exon 3 of ORF203 (lane 1), exon 1 of rps12' (lane 2), 5'-intron 1 of rps12' (lane 3), and the coding region of rpl20 (lane 4). P1 with an arrowhead indicates primary transcripts including ORF203 to rpl20; band a indicates processed mRNA containing rps12 (exon 1) and rpl20 genes; band b indicates processed mRNA for ORF203 with introns, and band b' indicates spliced mature mRNA of ORF203. Band m (open triangle) indicates mature mRNA for rps12 gene products. (B) Probes specific to 5'-intron 1 (lane 1), exon 2 (lane 2) of rps'12 gene, exon 1 (lane 3) and intron (lane 4) of ndh2 gene, and for the coding region (lane 5) of the trnL(CAA) gene. P2 and P2' with arrowheads indicate the two kinds of transcripts described in the text. P3 with an arrowhead indicates an additional transcript for ndh2. Band m with an open arrowhead indicates mature mRNA for the rps12 gene. (C) Probes specific to coding regions for exon 3 of the rps'12 and rps7 genes. Lane 1, probe for exon 3 of the rps12 gene; lane 2, for the rps7 gene. Band m with an open arrowhead indicates mature mRNA for the rps12 gene product.

detected (Fig. 2). The results obtained with the probes for the ORF203 (exon 3), the rps12 exon 1, the 5'-portion of rps12 intron 1, and the rpl20 gene are presented in Figure 2A. The rps12' was transcribed together with the upstream ORF203 and the downstream rpl20 as a primary RNA of 3.0 kb (band P1, Fig. 2A). We also detected a 1.55-kb RNA (band b) that was hybridized only with the ORF203 probe and a 1.45-kb RNA (band a) hybridized only with the rps12' and rpl20 probes (Fig. 2A). These two RNA species may be cleaved products of the 3.0-kb primary RNA transcript that was processed at a specific site between ORF203 and rps12'. Several bands were observed between bands P1 and a, which may be partially spliced molecules of primary transcripts (lane 2,

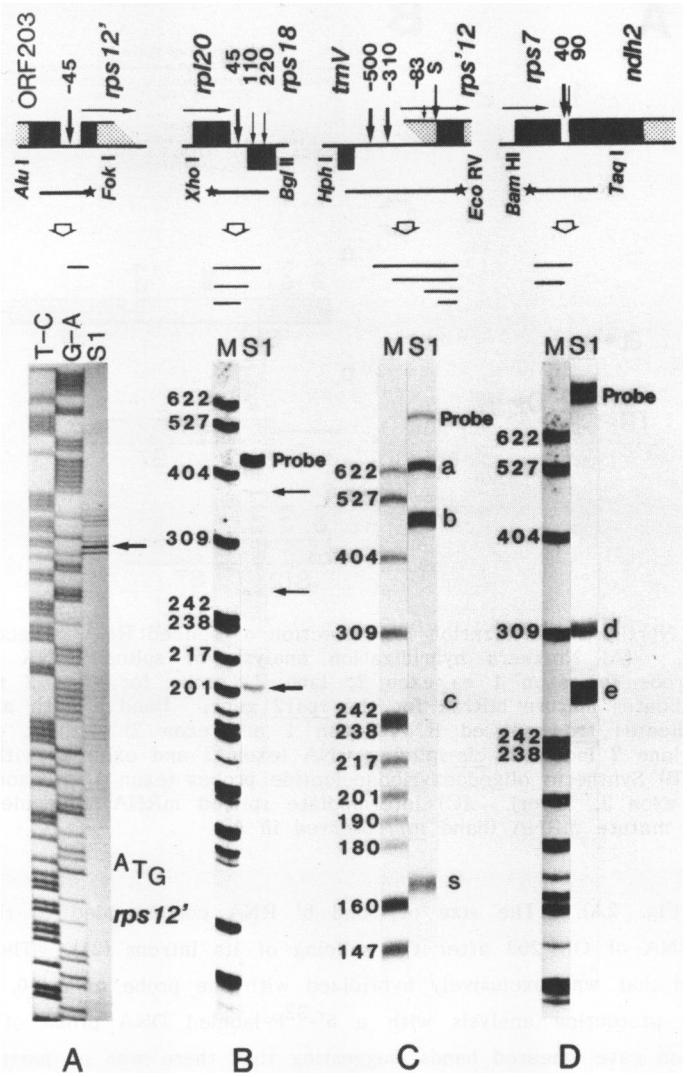


Figure 3. S1-nuclease protection analysis of transcripts. (A) DNA probe (AluI-FokI fragment) between the ORF203 (exon 3) and rps12' (exon 1) genes. (B) DNA probe (XhoI-BglII fragment) between the rpl20 and rps18 genes. (C) DNA probe (HphI-EcoRV fragment) covering the spacer region between the trnV(GAC) and rps'12 genes. (D) DNA probe (BamHI-TaqI fragment) covering the rps7 and ndh2 genes. Above each panel, signals from S1-nuclease protection analysis are shown with the gene organization. Numbers with arrows indicate the site of signals from the following exons or the coding region of the respective genes. Asterisks indicate the site of ^{32}P-labeled ends. Hatched boxes are introns.

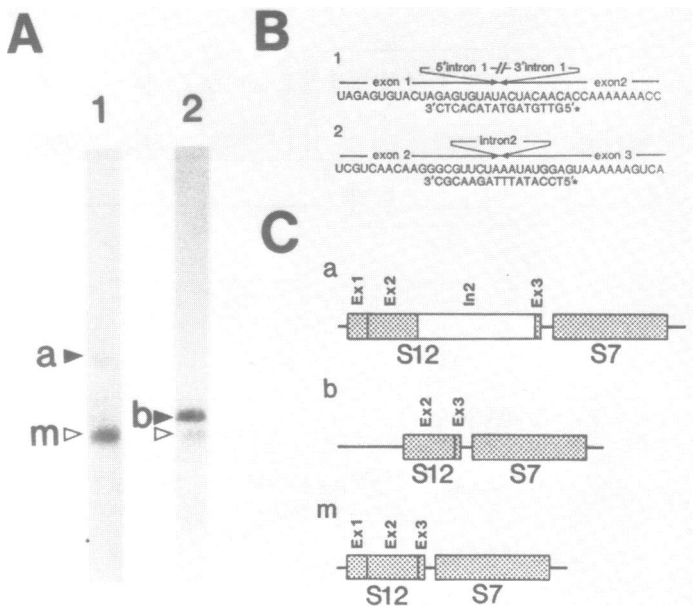


Figure 4. Northern hybridization for detection of spliced RNA molecules of the rps12 gene. (A) Northern hybridization analysis of spliced RNA transcripts. Lane 1, probe for exon 1 to exon 2; lane 2, probe for exon 2 to exon 3. Band m indicates mature mRNA for the rps12 gene. Band a with an arrow in lane 1 indicates trans-spliced RNA (exon 1 and exon 2) without cis-splicing. Band b in lane 2 indicates cis-splicing RNA (exon 2 and exon 3) without trans-splicing. (B) Synthetic oligodeoxyribonucleotide probes (exon 1 to exon 2, upper; exon 2 to exon 3, lower). (C) Intermediate spliced mRNA molecules (bands a and b) and mature mRNA (band m) observed in A.

3 and 4, Fig. 2A). The size of band b' RNA corresponded to that of the mature mRNA of ORF203 after the splicing of its introns (21). There was no unique band that was exclusively hybridized with the probe of rpl20. However, S1-nuclease protection analysis with a 5'-^{32}P-labeled DNA probe of the rpl20 coding region gave smeared bands, suggesting that there was no particular fixed end for the 5'-leader of rpl20 mRNA (data not shown).

Northern hybridization was done with a variety of probes for the 3'-portion of intron 1 (the 5'-leader sequence of exon 2), exon 2 of the rps12 gene, exon 1 of ndh2 gene, an intron of ndh2 gene, and the trnL(CAA) gene (Fig. 2B). The results showed that: (i) the primary transcripts (4.2 kb, band P2; 4.0 kb, band P2') correspond in their entire region from rps'12 (exons 2 and 3) to the trnL(CAA) gene; (ii) the major bands (1.9 kb, 1.4 kb, and 1.2 kb; bands c, d, and e, respectively; Fig. 2B) were mRNA molecules processed

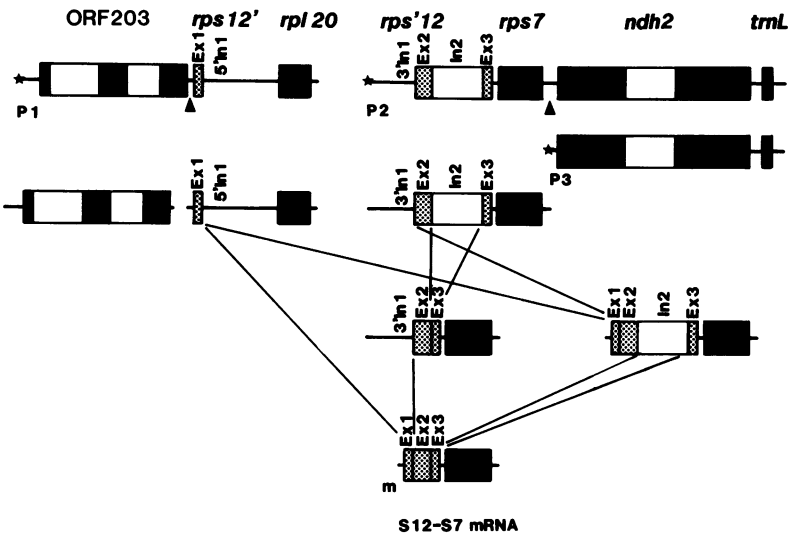


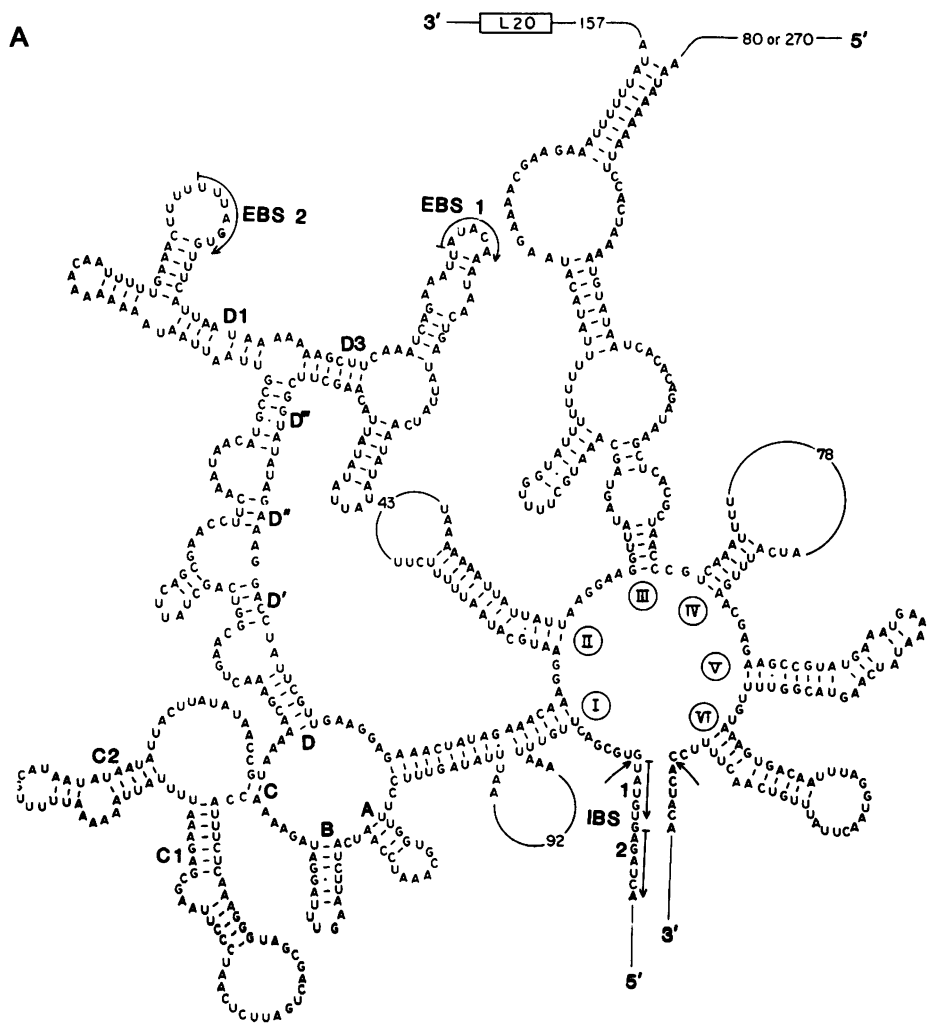
Figure 5. Schematic pathways of mRNA maturation. Arrowheads indicate major processing sites of transcripts. Asterisks indicate the 5'-terminals of primary transcripts. Dotted boxes indicate exons of the rps12 gene. In other genes, solid box are exons and white boxes are cis introns. Lines indicate the processes of trans- and cis-splicing.

upstream from the ndh2 gene; and (iii) there was an additional transcript for the ndh2 gene (band P3 in lanes 3 and 4, Fig. 2B). A large amount of trnL(CAA) gene product was found in the RNA preparation (lane 5, Fig. 2B). This may not reflect only the stability of tRNA molecules but also the presence of additional initiation sites for trnL(CAA) transcription alone. Northern hybridization with probes specific to the coding region for rps12 exon 3 and rps7 gave the same pattern, indicating that there was no RNA processing between them (lanes 1 and 2, Fig. 2C). The linkage of rps12 and rps7 was maintained on the mRNA, like that in E. coli (22).

Processing of two separate rps12 transcripts (exon 1 and exons 2-3).

S1-nuclease protection analysis at the 5'-flanking region of the rps12 exon 1 showed that the processing site was between the ORF203 and the rps12 exon 1 (Fig. 3A). This result coincided with Northern hybridization analysis, which showed that there were major processed mRNA molecules for the ORF203 gene (see band b, Fig. 2A) and for rps12'-rpl20 (see band a, Fig. 2A). The major termination site was 45 nucleotides downstream from the rpl20 gene, although two minor signals (110 nucleotides and 220 nucleotides downstream) were also observed (Fig. 3B). To identify the initiation site of transcripts in the gene

cluster of rps'12, S1-nuclease protection analysis was done with a DNA probe containing the HphI-EcoRV fragment, giving four bands (Fig. 3C). One could correspond a spliced RNA (band s, Fig. 3C). A minor band c (Fig. 3C) corresponded to the junction site (83 nucleotides upstream from the 5'-terminal of exon 2) of inverted repeat and large single copy regions, probably because of competitive hybridization with transcripts in the other inverted repeat region. This band, therefore, did not correspond to the initiation site of the transcription. Two additional major bands were detected: one corresponded to the RNA molecule that started 500 nucleotides upstream from the 5'-terminal



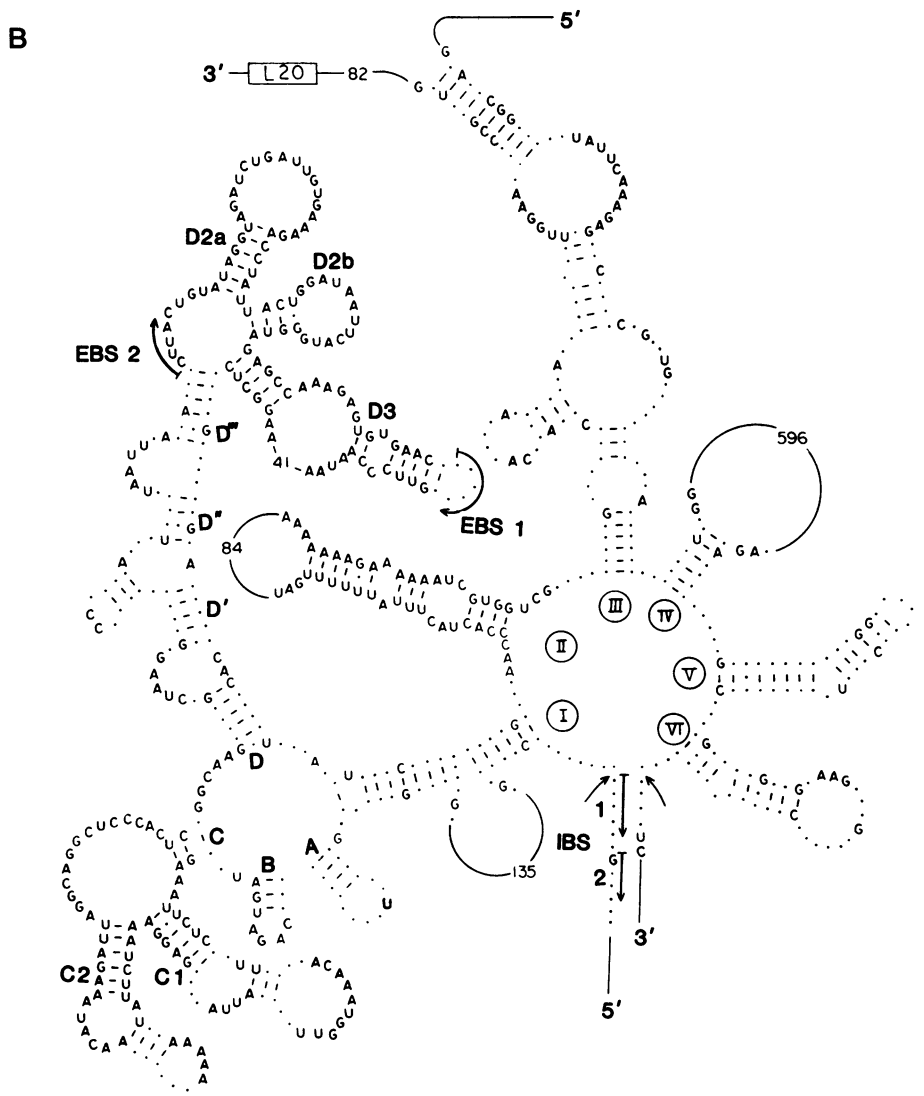


Figure 6. A model of secondary structure of trans-split introns in the rps12 gene. (A) Liverwort; (B) Tobacco (dots indicate the same nucleotides as liverwort). Arrows indicate splicing sites. Abbreviations IBS and EBS with arrows indicate the intron-binding and exon-binding sites (2, 23, 24). The nomenclature of stem-loop structure follows that of the model of Jacquier and Michel (23).

of exon 2 of rps12 (band a, Fig. 3C), and the other was 310 nucleotides upstream from exon 2 (band b, Fig. 3C). The former could be for the initiation of transcripts of the gene cluster, because of the presence of promoter-like sequences upstream ("TTGACC" and "TAAAAT" as "-35"- and "-10"- sequences, respectively). RNA molecules corresponding to these signals in size were found in the Northern hybridization analysis (see bands P2 and P2', Fig. 2B). S1-nuclease protection analysis to detect the processing site between the rps7 and ndh2 genes gave evidence of two major processed RNA molecules (40 nucleotides and 90 nucleotides downstream from the termination codon of the rps7 gene, bands d and e, Fig. 3D). Major processings between ORF203 and rps12', and between rps7 and ndh2 may be required for the folding of the two separate transcripts described below.

Independent trans- and cis-splicing in the rps12 gene.

To confirm the presence of trans- and cis-spliced mRNA molecules, synthetic oligodeoxyribonucleotides 16 nucleotides long consisting of 8 nucleotides complementary to the 3'-end of a 5'-exon and 8 nucleotides complementary to the 5'-end of the following 3'-exon were used as probes for Northern hybridization (Fig. 4B). Spliced RNAs of exon 1 and exon 2 were seen as two bands: the major band m was mature mRNA for the rps12 gene, and the minor band a corresponded to trans-spliced mRNA molecules, but it still carried intron 2 (lane 1, Fig. 4A). A probe for spliced RNA of exon 2 and exon 3 also gave two bands, with the major band b having the 3'-half of intron 1, and the minor band m corresponding to the mature RNA for the rps12 gene (lane 2, Fig. 4A). These results indicate the independent occurrence of trans- and cis-splicing in the transcripts of the rps12 gene (Fig. 4C), although cis-splicing was apparently more efficient than trans-splicing in the liverwort chloroplast.

DISCUSSION

The results obtained are summarized in Figure 5, which is a probable scheme for the mRNA maturation pathways in the chloroplasts. Note that in this analysis we detected a variety of precursor RNAs relatively abundant compared to the amount of matured mRNAs. This is mainly because we prepared the liverwort chloroplasts from the cell cultures in the late growth stage (2 weeks old, and grown under continuous illumination). If we transferred the cells to fresh medium, mature mRNAs in the chloroplasts became predominant within a day or two, suggesting the conversion of accumulated intermediate RNAs to mRNAs (22).

Figure 6A illustrates a folding model of intron 1 of the rps12 gene for its trans-splicing. On the identified precursor RNAs of rps12' and rps'12, there are stretches of nucleotide sequences complementary with each other as deduced from the DNA sequence, and the model was constructed by assuming base pairings between them. A simplified form of this model has been published by elsewhere (2). The overall structure has characteristics of group II introns, with its six stem-loop domains (5), but with an interruption in the third loop region. The exon- and intron-binding sites (the EBS and IBS sequence elements) are also indicated in the Figure; they are complementary with each other and essential for the splicing (23, 24). A structure comparable to the tobacco intron 1 of rps12 can be deduced from the published sequence (Fig. 6B). Zaita et al. (9) has reported the presence of complementary sequences ("transon 1" and "transon 2"). The sequence transon 1 corresponds to the regions from A to C in stem-loop I, and transon 2 is in the loop portion of stem-loop IV in our model. In the liverwort genome, a complementary structure could not be formed in the regions corresponding to the tobacco transons. Evidence for our model was the compensatory base substitutions observed between liverwort and tobacco to keep base pairings at the stem regions. Two mutations that disrupt the splicing of bI1 intron in yeast mitochondria have been mapped in the stem portion of stem III (25), indicating the importance of the stem structure for splicing. The intermolecular base pairings may participate in the folding of group II intron as a "ribozyme" and in the selection of an adequate partner molecule. There was no complementary sequence to the stem portion of stem-loop III of the trans-split intron in any other introns of the liverwort chloroplasts (1). Thus, the exon shuffling in chloroplast RNA splicing could not be taken into account. Self-splicing of chloroplast group II intron has not yet reported, but catalytic activity would essentially reside in the introns, depending on the tertiary folding structure. The resultant RNA complex may be essential for the removal of trans-split intron.

ACKNOWLEDGMENTS

This research was supported in part by a Grant-in-Aid for Special Research Projects from the Ministry of Education, Science, and Culture, Japan. The authors thank Dr. H. Fukuzawa, Institute for Applied Microbiology, University of Tokyo, for discussions.

* To whom correspondence should be addressed

REFERENCES

1. Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. (1986) Nature **322**, 572-574.
2. Ozeki,H., Ohyama,K., Inokuchi,H., Fukuzawa,H., Kohchi,T., Sano,T., Nakahigashi,K. and Umesono,K. (1987) Cold Spring Harbor Symp. Quant. Biol. **52**, 791-804.
3. Umesono,K and Ozeki,H. (1987) Trends in Genetics **3**, 281-287.
4. Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwonge,J., Obokata,K., Yamaguchi-Shinozaki,K., Ohto,C., Torazawa,K., Meng,B.Y., Sugita,M., Deno,H., Kamogashira,T., Yamada,K., Kusuda,J., Takaiwa,F., Kato,A., Tohdoh,N., Shimada,H. and Sugiura,M. (1986) EMBO J. **5**, 2043-2049.
5. Michel,F. and Dujon,B. (1983) EMBO J. **2**, 33-38.
6. Fukuzawa,H., Kohchi,T., Shirai,H., Ohyama,K., Umesono,K., Inokuchi,H. and Ozeki,H. (1986) FEBS Lett. **198**, 11-15.
7. Fromm,H., Edelman,M., Koller,B., Goloubinoff,P. and Galun,E. (1986) Nucl. Acids Res. **14**, 883-898.
8. Torazawa,K., Hayashida,N., Obokata,J., Shinozaki,K. and Sugiura,M. (1986) Nucl. Acids Res. **14**, 3143.
9. Zaita,N., Torazawa,K., Shinozaki,K. and Sugiura,M. (1987) FEBS Lett. **210**, 153-156.
10. Hildebrand,M., Hallick,R.B., Passavant,C.W. and Bourque,D.P. (1988) Proc. Natl. Acad. Sci. USA. **85**, 372-376.
11. Koller,B., Fromm,H., Galun,E. and Edelman,M. (1987) Cell **48**, 111-119.
12. Kück,U., Choquet,Y., Schneider,M., Dron,M. and Bennoun,P. (1987) EMBO J. **6**, 2185-2195.
13. Choquet,Y., Goldschmidt-Clermont,M., Girard-Bascou,J., Kück,U., Bennoun,P. and Rochaix,J. (1988) Cell **52**, 903-913.
14. Solnick,D. (1985) Cell **42**, 157-164.
15. Konarska,M.M., Padgett,R.A. and Sharp,P.A. (1985) Cell **42**, 165-171.
16. Jacquier,A. and Rosbash,M. (1987) Science **234**, 1099-1104.
17. Sharp,P.A. (1987) Cell **50**,147-148.
18. Ohyama,K., Wetter,R.A., Yamano,Y., Fukuzawa,H. and Komano,T. (1982) Agric. Biol. Chem. **46**, 237-242.
19. Kohchi,T., Yoshida,T., Komano,T. and Ohyama,K. (1988) EMBO J. **7**, 885-891.
20. Berk,A.J. and Sharp,P.A. (1977) Cell **12**, 721-732.
21. Kohchi,T., Ogura,Y., Umesono,K., Yamada,Y., Komano,T., Ozeki,H. and Ohyama,K. (1988) Curr. Genet. in press.
22. Post,L.E. and Nomura,M. (1980) J. Biol. Chem. **255**, 4660-4666.
23. Jacquier,A. and Michel,F. (1987) Cell **50**, 17-29.
24. Michel,F. and Jacquier, A (1987) Cold Spring Harbor Symp. Quant. Biol. **52**, 201-212.
25. Schmelzer,C., Schmidt,C., May,K. and Schweyen,R.J. (1983) EMBO J. **2**, 2047-2052.