ABSTRACT

The rps12 gene in tobacco chloroplasts consists of three exons that code for a polypeptide with homology to Escherichia coli ribosomal protein S12. C-terminal exons 2 and 3 of rps12 are located in the inverted repeat regions of the tobacco chloroplast genome. Exon 1 of rps12 is 29 kilobase pairs downstream of the nearest copy of exons 2 and 3 and 69 kilobase pairs away from the distal copy of exons 2 and 3. RNA gel blot hybridization analysis and primer extension sequencing of cDNA to rps12 encoding RNAs indicate that exons 1 and exons 2 and 3 are encoded on separate transcripts. Exon 1 and exons 2 and 3 are covalently ligated in the correct reading frame in rps12 mRNA. These results indicate that a bimolecular (trans-) splicing event occurs during the formation of mature rps12 mRNA.

The rps12 gene of tobacco chloroplasts encodes a chloroplast ribosomal protein that is homologous to Escherichia coli ribosomal protein S12. C-terminal exons 2 and 3 of rps12 (Fig. 1) are in the inverted repeat regions of the tobacco chloroplast genome (1, 2). Analysis of primary sequence (1, 3), and of possible secondary structure (2), suggested that an intron was present upstream of exon 2 (1, 2). Exon 1 of rps12 was identified (4) 29 kilobase pairs (kb) downstream from the nearest copy of exon 2, and 69 kb from the distal copy of exon 2 in the tobacco chloroplast genome (Fig. 1). Following codon 38 of rps12 exon 1 is a 5′-consensus intron splice boundary sequence (1, 3) of 5′ GTGCG 3′ (4).

Electron microscopic analysis of DNA:RNA hybrids identified several RNA species that hybridized to the rps12 exon 1 coding region and that have 5′ ends positioned at various distances upstream of exon 1 (5). Distinct RNA species hybridized to a cloned fragment from the rps12 exon 2–3 locus (5). Another class of RNA molecules contained exon 1 directly ligated to exon 2 (5). The splice junction between exon 1 and exon 2 has been determined by cDNA sequencing of rps12 mRNA (6). These results suggest that rps12 exon 1 and exons 2–3 are separately transcribed and that the separate transcripts undergo a bimolecular (trans-) splicing reaction during the formation of mature rps12 mRNA.

We present a characterization of the expression of the rps12 loci of Nicotiana tabacum chloroplast DNA. We show that exon 1 of rps12 is ligated, in the correct reading frame, to exon 2 of mature rps12 mRNA. RNA gel blot hybridization analysis reveals a complex mixture of RNA species formed during maturation of rps12 mRNA. Distinct RNA species encoding exon 1 and exons 2–3 have been identified by RNA gel blot analysis. Different starts of transcription for exon 1 and exon 2–3 have been identified. These results show that two transcripts from the rps12 gene of tobacco chloroplasts undergo a trans-splicing event during mRNA maturation and identify sequences that may be involved in initiation of rps12 transcription.

METHODS

Cloning and DNA Sequencing. By using standard techniques (7), subclones in M13mp18 or -mp19 vectors (8) were derived from clones of N. tabacum cv. Petite Havana chloroplast DNA (9). Restriction endonucleases and T4 DNA ligase were from Bethesda Research Laboratories or New England Nuclear. Alkaline phosphatase was from New England Nuclear. All enzymes were used under recommended conditions (7).

Labeling of Probes. Primer 2 (5′ CAACCTTTAGCTTAA 3′) and primer 8 (5′ CCGATCGATTGCGTAAAGCC 3′) were synthesized by Nancy Istock (Department of Molecular and Cellular Biology, University of Arizona). Primers were 5′-end-labeled in a reaction mixture (25-μL) consisting of 20 pmol of primer, 40 pmol of [32P]ATP (7000 Ci/mmol; 1 Ci = 37 GBq; New England Nuclear), 5.6 units of T4 polynucleotide kinase [International Biotechnologies (New Haven, CT) or Bethesda Research Laboratories], 50 mM Tris-HCl (pH 7.5), 50 mM MgCl2, and 1 mM dithiothreitol. The mixture was incubated at 37°C for 30 min, then Na2EDTA (to 20 mM) and 5 μg of carrier tRNA were added. Nucleic acids were ethanol precipitated, pelleted, and washed twice with 70% (vol/vol) ethanol. Samples were vacuum dried and resuspended in 50 μL of water.

Hybridization probes were derived from M13 subclones (see Fig. 4A) by synthesizing a labeled ([α-32P]dATP, 3000 Ci/mmol, New England Nuclear) complementary strand (8) with the Klenow fragment of DNA polymerase I (Bethesda Research Laboratories).

RNA Isolation. Young tobacco leaves (N. tabacum cv. Petite Havana) were homogenized (130-g batches) in 260 ml of ice-cold grinding buffer (GB; 50 mM Tris-HCl, pH 8.0/350 mM mannitol/7 mM Na2EDTA/5 mM 2-mercaptoethanol/0.1% bovine serum albumin/1 mM aurinic acid) and remained at room temperature for 10 min. The lysed chloroplasts were extracted twice with water-saturated phenol and twice with ethyl ether, followed by ethanol precipitation of nucleic acids. Nucleic acids were precipitated, washed twice with 70% (vol/vol) ethanol, vacuum dried, and resuspended at 2 mg of nucleic acid per mL in 1 mM aurinic acid (in diethylpyrocarbonate-treated water). One-tenth volume of 10× medium salt restriction buffer (7) and RNase-free DNase at 18 ng/μL (final concentration) (Promega Biotec, Madison, WI) was added. After 1 hr at 37°C, the solution was phenol extracted twice, ether extracted twice, and ethanol precipitated. The RNA was pelleted, washed twice with 70% (vol/vol) ethanol, vacuum dried, and resuspended in diethylpyrocarbonate-treated water.

*To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Fig. 1.  Map of the rps12 loci on the tobacco chloroplast genome. (A) Linearized map of the coding regions (solid bars) of rps12 exon 1, open reading frame (ORF) 73, and rpl20. (B) The circular tobacco chloroplast genome showing inverted repeats (blocks concentric to the main circle) containing 16S and 23S rRNA genes (solid bars) and the location and direction of transcription (arrows) of the coding loci of rps12. (C) Linearized map of the region in the inverted repeat encoding rps12 exons 2 and 3, rps7, and ORF 119 (solid bars).

RNA (1 mg per gradient) was size-fractionated on 11.6-ml 5–25% sucrose gradients (in 100 mM NaCl/10 mM Tris-HCl, pH 7.4/1 mM Na₂EDTA). Centrifugation was at 22°C in an SW 41 rotor for 16 hr at 28,000 rpm. Aurincarboxylic acid was removed from RNA by chromatography on a 6.5-ml Sephadex G-25 (Pharmacia) column. The elution buffer was 0.1 M NaOAc, pH 5.2/0.1% NaDodSO₄.

**Primer-Extension Sequencing of cDNA.**  ³²P-end-labeled primer (0.6 pmol) was annealed to 3 µl of size-fractionated RNA in annealing buffer (AB; 50 mM Tris-HCl, pH 8.3/60 mM NaCl/10 mM dithiothreitol) in a total reaction volume of 12 µl. This annealing mixture was heated to 65°C for 5 min, placed at 42°C for 3 hr, slowly cooled (~1 hr) to 35°C, and placed on ice. To a 2-µl aliquot of the annealing mixture was added 2 µl of reaction solution [0.4 mM dATP, 0.4 mM dGTP, 0.4 mM dCTP, and 0.4 µM dGTP, 50 mM Tris-HCl (pH 8.3), 60 mM NaCl, 10 mM dithiothreitol, 30 mM magnesium acetate, and avian myeloblastosis virus reverse transcriptase at 0.25 unit/µl (Life Sciences, St. Petersburg, FL)], and 1 µl of the appropriate dideoxynucleotide triphosphate (80 µM in AB) or 1 µl of AB for the no dideoxynucleotide control sequencing (N) reaction. After 15 min at 37°C, 1 µl of chase solution (reverse transcriptase at 0.25 unit/µl of AB) was added. After an additional 15 min at 37°C, 6 µl of stop solution [95% (vol/vol) formamide, 10 mM Na₂EDTA, 0.1% xylene cyanol FF, and 0.1% bromophenol blue] was added. Samples were boiled for 5 min prior to loading 3 µl per lane on a denaturing 6% polyacrylamide gel. After electrophoresis the gel was transferred to a sheet of Whatman 3MM paper, wrapped in SaranWrap, and autoradiographed at −80°C, using Kodak XAR-5 film.

Fig. 2. Primer-extension sequencing of trans-spliced rps12 mRNA. The 5'-end-labeled primer 2, complementary to rps12 exon 2 RNA, was hybridized to mature rps12 mRNA, and dideoxy sequencing was performed. The autoradiogram displays the products of the A, T, C, G, and no dideoxynucleotide control (N) sequencing reactions in lanes labeled accordingly. The cDNA sequence is listed at left, beginning with the initiation codon. The vertical arrow denotes the exon 1–2 splice junction.
RNA Gel Blot Hybridizations. Chloroplast RNA (15 μg per lane) was electrophoresed on methylene-pyridine hydroxide gels (10) and transferred to Schleicher & Schuell PH79 (0.1-μm) nitrocellulose. Prehybridization was at 42°C for 24 hr in a solution of 50% (vol/vol) formamide, 1× Denhardt's solution (1× Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin) in 3× SSC (1× SSC is 150 mM NaCl/15 mM sodium citrate, pH 7), 0.1% NaDodSO₄, 6× NET (1× NET is 150 mM NaCl/1 mM Na₂EDTA/15 mM Tris-HCl, pH 8.0), and degraded herring sperm DNA at 100 μg/ml (Sigma). Hybridizations were at 42°C for 24 hr in a solution of 50% (vol/vol) formamide, 2× Denhardt's solution, 0.1% NaDodSO₄, 6× NET, degraded herring sperm DNA at 100 μg/ml, and probe at 3 ng/ml. Filters were washed twice for 15 min at room temperature in 2× SSC, 0.1% NaDodSO₄, and twice at room temperature in 0.1× SSC/0.1% NaDodSO₄. The filters were autoradiographed at -80°C, using Kodak XAR-5 film.

RESULTS

In-Frame Ligation of Exons 1 and 2 in rps12 mRNA. Primer-extension dieoxy sequencing of size-fractionated chloroplast RNA revealed that exon 1 and exon 2 are ligated in mature rps12 mRNA (Fig. 2). The correct reading frame is maintained across the exon 1–2 splice junction, in agreement with results of Zaita et al. (6).

5’ Ends of RNAs Encoding rps12 Exon 1 and Exon 2–3/rps7. Exons 2 and 3 of rps12 and rps7 are cotranscribed (2). Separate 5′ ends of RNAs encoding rps12 exon 1 and exon 2–3/rps7 were identified by primer extension cDNA sequencing (Fig. 3). The 5′ end of an rps12 mRNA species was located 52 nucleotides upstream of the initiation codon for exon 1 (Fig. 3 Upper Left). This result locates the transcription start most likely nearest exon 1 (5). The start of transcription for rps12 exon 2–3/rps7 was identified 1241 or 1242 nucleotides upstream of the 5′ end of rps12 exon 2 (Fig. 3 Upper Right). The length of the exon 2–3/rps7 primary transcript is ~2800 nucleotides, based on our determination of the 5′ end of this transcript and the heteroduplex analysis of Fromm et al. (2).

The 5′ end of the rps12 mRNA species begins downstream of typical (11, 12) chloroplast consensus promoter “−35” and “−10” sequences (Fig. 3 Lower). The sequence of the −10 region of the exon 2–3/rps7 primary transcript is very similar to that upstream of exon 1 (Fig. 3 Lower). However, a possible −35 sequence for the exon 2–3/rps7 transcript has less homology and an altered location relative to other chloroplast −35 sequences (11, 12).

RNA Gel Blot Analysis of rps12 Transcripts and Processing Intermediates. Transcripts encoding rps12 exon 1, and rps12 exon 2–3/rps7, and several RNA species that arise during their processing have been identified by RNA gel blot analysis (Fig. 4). Bands a and b (>3500 nucleotides) probably correspond to the highest molecular weight transcripts.
from the exon 1 coding region (5). Band g (3000 nucleotides long), which hybridizes to rps12 exon 2 and 3, rps12 intron 1 and 2, and rps7 probes, corresponds to the exon 2-3/rps7 primary transcript. Band h (2400 nucleotides) corresponds to a transcript of rps12 exon 2-3/rps7 from which intron 2 has been excised. Bands c and d (specific exon 1 probe) are of similar molecular weight to bands g and h. Since RNA species containing both exons 1 and 2 separated by intervening sequences were not observed by Koller et al. (5), bands c and d probably correspond to transcripts from the exon 1 coding region. Band e (1500 nucleotides) hybridizes to all three rps12 exons and rps7 but not to the intron-specific probes, as expected for a mature dicistronic rps12/rps7 mRNA. This has been confirmed by cDNA sequence analysis of size-fractionated RNA (Fig. 2 and unpublished results). Band f (exon 1 probe) represents the lowest molecular mass (1260 nucleotides) RNA species detected containing exon 1. The 5' end of this species probably corresponds to the 5' end of an rps12 mRNA determined by cDNA sequencing (Fig. 3 Upper Left). Band k may correspond to a processing product of the rps12 exon 2-3/rps7 transcript (5). Band l (specific intron 1 probe) corresponds roughly in length (1170 nucleotides) to the sequence between exon 2 and the 5' end of the rps12 exon 2-3/rps7 primary transcript. Band i is unique to the intron 2 specific probe (2-55). Band j (intron 2 probe) migrated at ~1620 nucleotides. No RNA of molecular weight corresponding to excised linear intron 2 (536 nucleotides) was detected. Further work is required to determine if band j might be a circular or lariat form of excised intron 2.

**DISCUSSION**

Evidence for Trans-Splicing of rps12 Transcripts. We have determined that the rps12 gene is expressed as mature mRNA (Fig. 4, band e) in tobacco chloroplasts, with exon 1 ligated to exon 2 in the correct reading frame (Fig. 2). Mature rps12 mRNA could result from the processing of a 127-kilobase primary transcript, originating upstream of exon 1 and proceeding through the nearest copy (same DNA strand) of exon 2-3/rps7. However, many genes that would be encompassed by such a transcript have defined transcript sizes consistent with the location of their putative promoters.

**Fig. 4.** RNA gel blot hybridization analysis of rps12 transcripts. (A) Diagram of the coding loci of rps12 exon 1 (Upper) and rps12 exon 2-3/rps7 (Lower) and the location (brackets) of the single-stranded M13 clones used as probes. The vertical arrows position the start sites for transcription for exon 1 (Upper) and exon 2-3/rps7 (Lower) as determined by the data in Fig. 3. The exon 1 probe contains a 170-bp Sau3AI fragment extending 44 bp upstream and 14 bp downstream of rps12 exon 1. Fragment Bam24 (9) extends from 104 bp upstream of rps12 exon 2 to 270 bp downstream of the initiation codon of rps7. Fragment Bam28 (9) is a 426-bp BamHI fragment containing sequences entirely within rps12 intron 1. Clone 1-5 contains a 315-bp Sau3AI fragment extending from 104 bp upstream of rps12 exon 2 to 21 bp upstream of the 3' end of rps12 exon 2. Clone 2-55 contains a 297-bp Sau3AI fragment having sequences entirely within rps12 intron 2. Clone 4-23 contains a 133-bp Sau3AI fragment extending from 12 bp upstream of rps12 exon 3 to 38 bp downstream of the rps7 initiation codon. (B) Autoradiograms of the RNA gel blots probed with the various fragments described in A. Probes are listed at the top of each autoradiogram. The letters a-i identify the RNA species detected by RNA gel blot hybridization.
and terminators (13), suggesting that transcription would not proceed through them. Distinct transcriptional starts for exon 1 and exon 2-3/rps7 were identified by cDNA sequencing. An RNA species corresponding to the rps12 exon 2-3/rps7 primary transcript was identified (Fig. 4, band g). RNA species of distinct size from the exon 1 coding region have been identified by RNA gel blot (Fig. 4, bands a–d and f) and heteroduplex (5) analyses. These results are consistent with separate transcriptional events occurring for exon 1 and exons 2-3. Therefore, we conclude that the expression of the tobacco chloroplast rps12 gene involves (i) separate transcriptional events for exon 1 and exons 2-3/rps7, (ii) a trans-splicing event, and (iii) other RNA processing to form mature rps12 mRNA.

Transcription Initiation and/or Processing Sites of rps12 Encoding RNAs. The 5’ end of an RNA species encoding rps12 exon 1 (Fig. 3) begins downstream of a consensus chloroplast promoter sequence (11, 12). Several other exon 1-containing RNA species (Fig. 4, bands a–d) were observed that may have 5’ ends upstream of this site, as suggested by heteroduplex analysis (5). It is unclear whether these RNA species arise from separate transcriptional events or from processing of larger transcripts. Several processing and/or transcription initiation sites have also been identified in the rps12 exon 2-3/rps7 region (5).

Different putative promoter sequences are located upstream of the initiation sites for transcription of the exon 1 and exon 2-3/rps7 coding loci (Fig. 3). Regulation of the levels of rps12 transcripts could be achieved through different promoter strengths for the two parts of the gene. Direct measurements of promoter strengths and transcript levels are required to test this hypothesis.

Trans-Splicing as a Possible Means of Gene Evolution. The exon-shuffling model of gene evolution proposes that gene segments encoding structural domains of a protein (exons) are moved by genomic rearrangements into other genes (14, 15). The results presented here, showing the involvement of two separate transcripts in the expression of a single gene, raise the question of whether genes originally consisted of units encoding a single transcript. In an evolutionary sense, structural domains of a protein could have been encoded by separate transcripts that then became trans-spliced to form a single message (16). Eventual reverse transcription of the trans-spliced or mature RNA into DNA, followed by insertion into the genome (17, 18), could result in the two parts of a gene becoming regulated under the same transcriptional control. Trans-splicing of RNA encoding rps12 exon 1 with transcripts from other genes could allow for chloroplast polypeptide variability; however, the complexity of the RNA gel blot hybridization data (Fig. 4) precludes determination of whether this occurs.

The results presented here document an example of a biologically occurring trans-splicing event involving the translated regions of a gene. The RNA gel blot hybridization results (Fig. 4) and other data (5) indicate that processing of rps12 transcripts produces a complex mixture of RNA species. Further characterization of the molecular species involved in rps12 transcription and processing is necessary to permit elucidation of the mechanism of trans-splicing and the maturation of rps12 mRNA.

The authors thank Barb Gandy, Nina Parmee, Steve Chipowski, and Xiao-lan Yang for technical assistance. We are especially indebted to Dr. Hans J. Bohnert for stimulating discussions and encouragement during the course of this work. This work was supported by Grant GM26937 from the National Institutes of Health to D.P.B. and Grants GM35659 and GM35659 from the National Institutes of Health to R.B.H. M.H. was supported during part of this work by an International Harvester Dissertation Research Fellowship. This publication no. 4453 of the Arizona Agricultural Experiment Station.