The EMBO Journal vol.10 no.2 pp.433-442, 1991

Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene

Donald W.Copertino[1] and Richard B.Hallick[1,2]

[1]Department of Molecular and Cellular Biology and [2]Department of Biochemistry,University of Arizona,Tucson, AZ 85721, USA

Communicated by J.-D.Rochaix

The *psbF* gene of chloroplast DNAs encodes the β-subunit of cytochrome b-559 of the photosystem II reaction center. The *psbF* locus of *Euglena gracilis* chloroplast DNA has an unusual 1042 nt group II intron that appears to be formed from the insertion of one group II intron into structural domain V of a second group II intron. Using both direct primer extension cDNA sequencing and cDNA cloning and sequencing, we have determined that a 618 nt internal intron is first excised from the 1042 nt intron of *psbF* pre-mRNA, resulting in a partially spliced pre-mRNA containing a 424 nt group II intron with a spliced domain V. The 424 nt intron is then removed to yield the mature *psbF* mRNA. Therefore, the 1042 nt intron of *psbF* is a group II intron within another group II intron. We use the term 'twintron' to define this new type of genetic element. Intermediates in the splicing pathway were detected by northern hybridization. Splicing of both the internal and external introns occurs via lariat intermediates. Twintron splicing was found to proceed by a sequential pathway, the internal intron being removed prior to the excision of the external intron. A possible mechanism for twintron formation by intron transposition is discussed.

Key words: chloroplast/*Euglena*/group II intron/intron transposition/twintron

Introduction

The introns in fungal and plant organelles have been classified into two groups based on several conserved sequences, and conserved core secondary structures (Michel *et al.*, 1982; Michel and Dujon, 1983). Group I intron excision occurs via a transesterification reaction that is initiated by an exogenous guanosine or guanosine nucleotide (Cech, 1986; Perlman *et al.*, 1990). The excised intron may also convert to a circular form by another transesterification reaction. Many of the group I introns self-splice *in vitro*. Intron secondary structure mediates this process (Bass and Cech, 1986; Cech, 1988). Group II introns have conserved core secondary structures and intron boundary sequences distinct from that of the group I introns (Michel *et al.*, 1982; Michel and Dujon, 1983; Keller and Michel, 1985). The secondary structure is represented as a central core with six radiating helical domains (I–VI) (reviewed in Michel *et al.*, 1989). Splicing, as determined by *in vitro* studies with self-splicing group II introns, occurs by a pair of transesterification reactions (Peebles *et al.*, 1986; Van der Veen *et al.*, 1986; Perlman *et al.*, 1990). The introns are excised via lariat intermediates in which a 2′ hydroxyl of an adenosine residue within domain VI initiates the nucleophilic attack at the 5′ splice junction. Domain V of the conserved secondary structure is absolutely required for the first transesterification reaction, and may also deliver the branch site of domain VI to the 5′ splice junction (Jarrell *et al.*, 1988). The similarity of the intron boundary sequences and the splicing mechanisms between group II introns and nuclear introns implies an evolutionary relationship between these introns (Cech, 1986; Perlman *et al.*, 1990).

The photosystem II reaction center core from chloroplast thylakoid membranes was reported to consist of four chloroplast DNA-encoded polypeptides, the D1 and D2-reaction center polypeptides and the α- and β-subunits of cytochrome b-559 (Nanba and Satoh, 1987). However a fifth component, the 4.5 kd product of the *psbI* gene, also copurifies with the photosystem II reaction center core (Webber *et al.*, 1989). The α- and β-subunits of apo-cytochrome b-559 are 9.3 and 4.4 kd, respectively (Widger *et al.*, 1985). Both subunits are oriented with the NH_2-terminus on the stromal side of the photosynthetic membranes, and the COOH-terminus on the lumenal surface (Tae *et al.*, 1988; Tae and Cramer, 1989). The specific function of cytochrome b-559 in the reaction center is unknown.

The genes for the cytochrome b-559 α- and β-subunits were mapped to a single locus in spinach chloroplast DNA (Westhoff *et al.*, 1985) and sequenced (Herrmann *et al.*, 1984). The *psbE* (for gene nomenclature, see Hallick, 1989) encodes the 83-residue α-subunit. Immediately downstream, and within the same operon, are the *psbF* gene, which encodes the 39-residue β-subunit, the *psbL* gene for a 38 residue photosystem II polypeptide (Webber *et al.*, 1989; Ikeuchi *et al.*, 1989), and an ORF of 40–42 residues for a yet unidentified polypeptide, likely from photosystem II. The *psbE-psbF-psbL-ORF40/42* genes from wheat (Hird *et al.*, 1986), *Oenothera hookeri* (Carillo *et al.*, 1986), *Marchantia polymorpha* (Ohyama *et al.*, 1986), tobacco (Shinozaki *et al.*, 1986), *Cyanophora paradoxa* (Cantrell and Bryant, 1987), *Euglena gracilis* (Cushman *et al*, 1988a), *Synechocystis* 6803 (Pakrasi *et al.*, 1988), rye (Kolosov *et al.*, 1989), barley (Chakhmakhcheva *et al.*, 1989), and maize (Haley and Bogorad, 1990) have been sequenced. In this diverse group of photosynthetic organisms, both the gene organization and the predicted amino acid sequences of the four polypeptides have been highly conserved.

The *Euglena* chloroplast *psbE-psbF-psbL-ORF 40/42* locus is exceptional with respect to the other chloroplast, cyanelle and cyanobacterial operons in two respects. First, In *Euglena* these 4 genes are co-transcribed with the *psaA* and *psaB* loci encoding the reaction center polypeptides of photosystem I as a *psaA-psaB-psbE-psbF-psbL-ORF42* operon. This juxtaposition of photosystem I and II genes in the same operon is unique to date. Second, this hexacistronic operon contains at least 12 group II introns, 3 in *psaA*, 6 in *psaB*,

2 in *psbE*, and 1 or more in *psbF* (Cushman *et al.*, 1988a and 1988b). Other known cytochrome b-559 genes lack introns. Of the 47 characterized group II introns of *Euglena* chloroplast DNA, 46 fall into the size range of 277−618 nt. The exception is the *psbF* intron, which has a length of 1042 nt (Cushman *et al.*, 1988a). In their excellent review on group II introns of mitochondria and chloroplasts, Michel *et al.* (1989) noted that the *psbF* intron could be interpreted as one group II intron inserted into another group II intron. Using both direct primer extension cDNA sequencing and cDNA cloning and sequencing, we have determined that the *psbF* intron is an intron-within-an-intron as predicted, which we are designating 'twintron'. A 618 nt internal intron is first excised from *psbF* pre-mRNA, resulting in a partially spliced pre-mRNA containing a 424 nt group II intron that is spliced in domain V. The 424 nt intron is then removed to yield the mature *psbF* mRNA.

Results

cDNA sequence analysis of partially-spliced twintron

The 1042 nt intron of the *Euglena* chloroplast *psbF* gene can be interpreted as two group II introns, one internal to the other (Figures 1 and 2). The external, or F1, intron is 424 nt in length with the 6 helical domains I−VI typical of group II introns (Michel *et al.*, 1989). It is adjacent to *psbF* exon 1 at its 5′ end, and *psbF* exon 2 at its 3′ end, with normal group II 5′ and 3′ boundary sequences. This intron is split into two halves by an internal group II, or F2, intron of 618 nt. As shown in Figure 2, the proposed insertion site for the internal intron is within domain V of the external intron. The complete sequence of this region is given in Cushman *et al.* (1988a; the proposed internal intron has coordinates 1346−1963 in Figure 2).

To initially test if the postulated twin intron structure is correct, cDNA sequence analysis was employed. This experiment was designed to detect an RNA processing intermediate with the internal F2 intron removed, and the split halves of domain V of intron F1 spliced. A synthetic, 20-mer of sequence 5′-ATTATTTCCTATAAAACTGC-3′, complementary to RNA spanning domains V and VI of intron F1, was used for direct, primer extension cDNA sequence analysis across the predicted twin intron splice junction. Purified chloroplast RNA was used as template. The location of the cDNA primer is highlighted in Figure 3B. The results are shown in Figure 3A.

With total chloroplast RNA as template, two sequence ladders can be clearly distinguished. The most prominent ladder is the cDNA sequence of the spliced F1 intron shown in upper case letters. The cDNA sequence begins 5′-CGAG... adjacent to the priming site, and continues ...ATACTTTCATA... As shown in Figure 3B, this is the cDNA sequence of spliced domain V of intron F1. This is the first direct evidence for an intron-within-an-intron. We have named this twin intron a 'twintron.' A minor cDNA sequence ladder can also be distinguished beginning after the 5′-CGAG... and continuing ...gaaagaaatag... As shown in Figure 3C, this is the cDNA sequence of the unspliced twintron. The presence of two distinct cDNA ladders is the expected result if the chloroplast RNA template is a mixture of unspliced and partially spliced *psbF* pre-mRNA. In support of this interpretation is the observation that several steps in the sequence ladder have only single bands, including

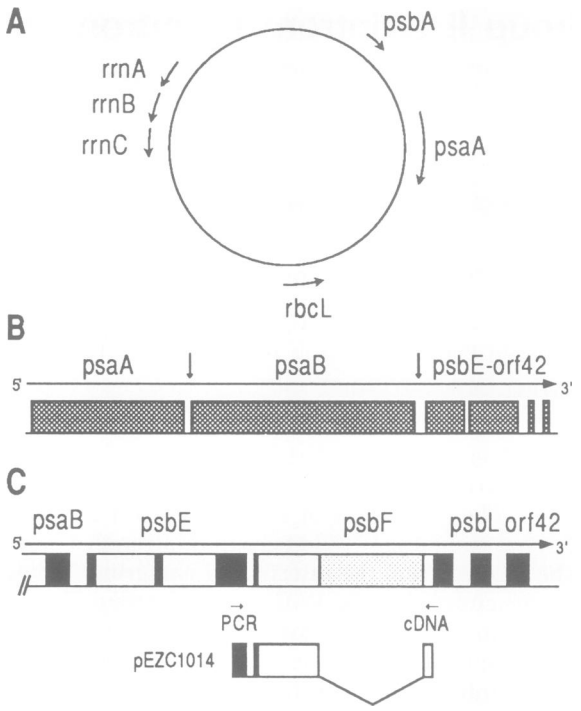

Fig.1. The *Euglena* chloroplast apocytochrome b-559 β-subunit gene *psbF* is encoded in the *psaA* operon. (**A**) Physical location of the *psaA* operon on the 145 kb *Euglena* chloroplast genome relative to other operons. The arrows indicate the direction of transcription. (**B**) The structure of the 12 kb *psaA* operon. The vertical arrows indicate the putative sites for endonucleolytic cleavage of the polycistronic transcript. (**C**) The organization of the *psbE-psbF-psbL-ORF42* (2.6 kb) genes and the experimental design for the analysis of the *psbF* twintron. The filled boxes correspond to exons and the open boxes correspond to introns. The stippled box within *psbF* denotes the internal intron F2. The locations of the cDNA and PCR primers used for the analysis of the *psbF* twintron are indicated. The cDNA insert of the plasmid pEZC1014 corresponding to the partially spliced twintron is also shown.

the region (C or t)ATAAA(G or a).. These single bands occur in regions where the unspliced and partially spliced twintron have the same sequence equidistant from the priming site.

One other feature of the cDNA sequence ladder of Figure 3A is noteworthy. There is a noticeable stop in the ladder with the sequence 5′-...CGAGGAAAGAA-OH, even in the no dideoxynucleotide control. As shown by an asterisk in Figure 3C, this stop corresponds to the expected position for branch formation in domain VI of an excised intron F2. If this interpretation is correct, the template for this cDNA must have been an intron lariat intermediate still connected to downstream sequences. The downstream sequences in this case was intron F1, which contained the sequence complementary to the cDNA primer.

PCR amplification of pre-mRNA and partially spliced mRNA

One limitation of direct cDNA sequencing is that the chloroplast RNA used as template contains a mixture of species complementary to the cDNA oligonucleotide primer. It could not be determined if the template for identification of the spliced F1 intron was the excised intron itself, or the partially spliced pre-mRNA. Likewise, the template for cDNA analysis of unspliced twintron could be unspliced pre-mRNA, or an excised twintron, that is subsequently spliced.

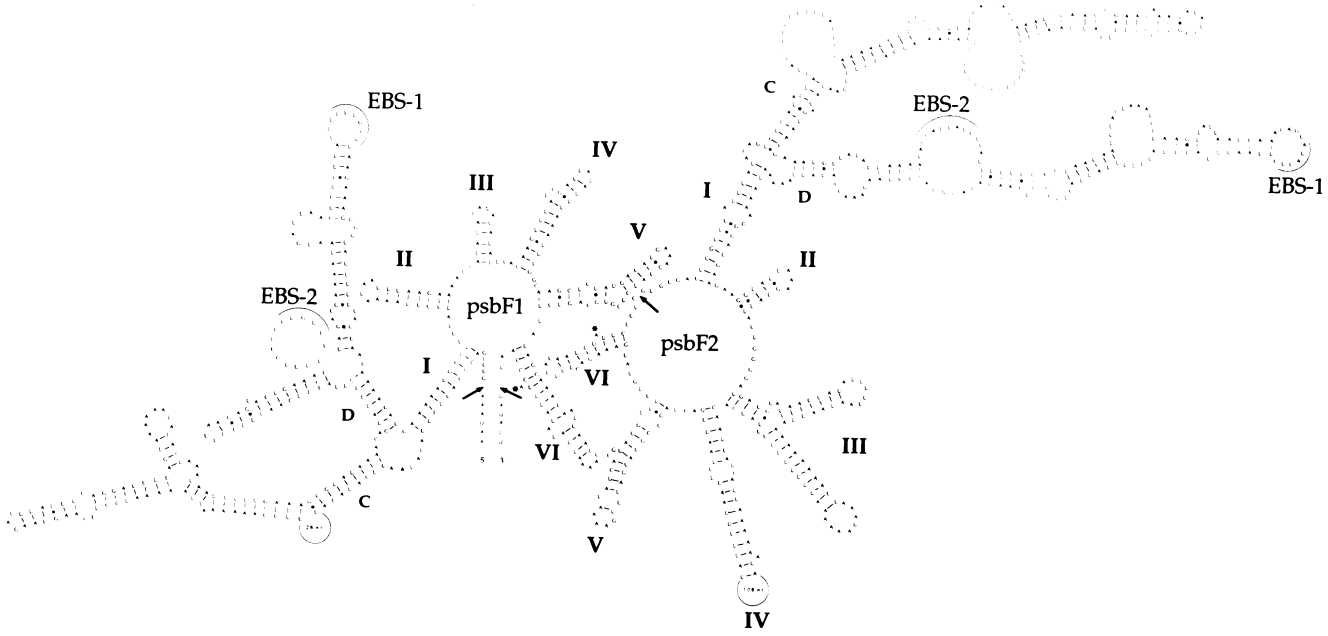

Fig.2. Secondary structural model for the *psbF* twintron. The external intron psbF1 is on the left and the internal intron psbF2 is on the right. The proposed twintron junction site is within domain V of the external intron F1 and is marked with an arrow. The domains I–VI for each intron are indicated in bold. The asterisks denote the adenine residues that are the proposed branch sites involved in lariat formation. Arrows also denote *psbF* exon and twintron boundaries. The secondary structural model is based on the model proposed by Michel *et al.*, 1989 and structural elements with known or suspected importance are indicated.

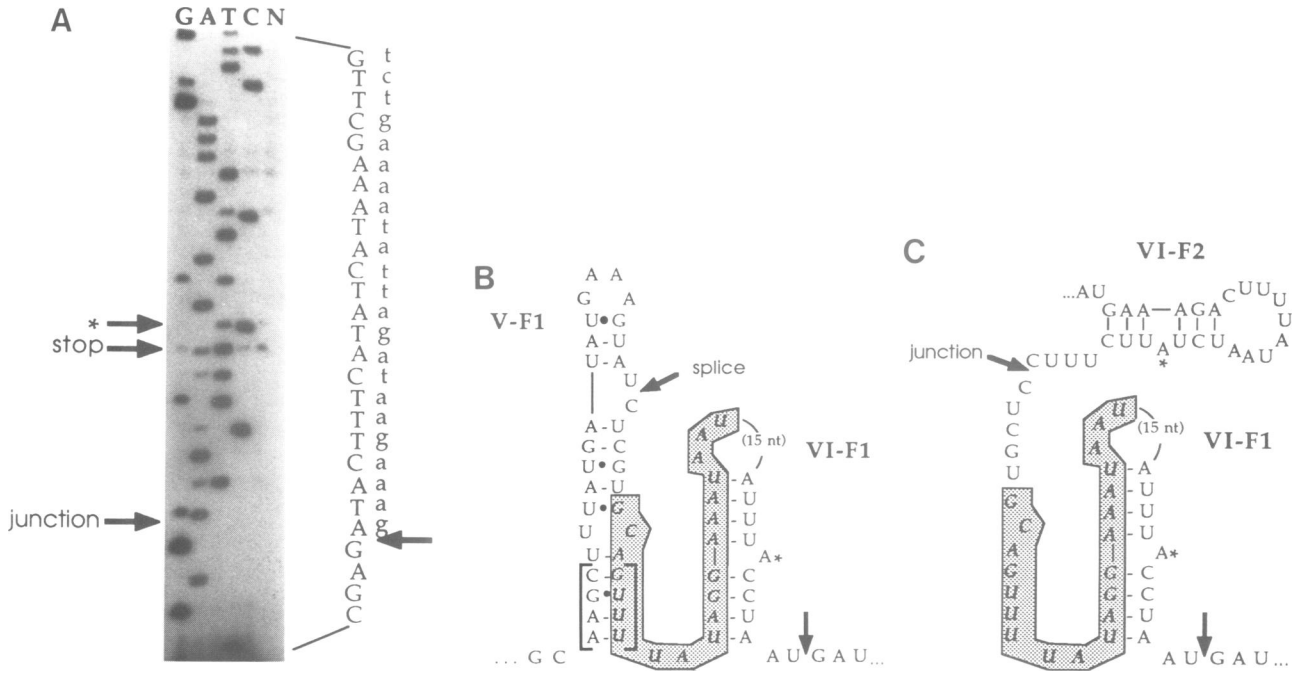

Fig. 3. The internal intron F2 is excised from the *psbF* twintron. (**A**) Primer extension cDNA sequence analysis of the partially spliced and unspliced *psbF* twintron. Total chloroplast RNA was used as template for dideoxy (dd) cDNA sequencing using AMV reverse transcriptase and a ^{32}P end-labeled oligonucleotide primer complementary to the region highlighted in Figure 3B. Primer extension reactions were done in the presence of either ddGTP (G), ddATP (A), ddTTP (T), ddCTP (C) or no dideoxynucleotides (N) as a control. The sequence complementary to the RNA is shown. The splice site of the F1 intron is indicated by the lower arrow at the left of the gel. The upper case letters at the right of the gel correspond to the major bands of the spliced F1 intron. The lower case letters correspond to the minor bands of the unspliced *psbF* twintron. The stop corresponding to the putative branch site of the F2 intron lariat intermediate is indicated by the higher arrow at the left of the gel. (**B**) Secondary structure of domain V of the F1 intron after splicing of the F2 intron. This RNA corresponds to the upper case letters in (**A**). The splice site of the F1 intron is indicated by the arrow. (**C**) Secondary structures of domain VI of the F1 and F2 introns prior to excision of the F2 intron. This RNA corresponds to the lower case letters in (**A**). The junction of the F1 and F2 introns in the unspliced twintron is indicated by the arrow. The bulging adenine residue marked with an asterisk in domain VI of F2 corresponds to the putative branch site of the F2 intron intermediate identified as a stop in the cDNA sequence ladder in (**A**). The location of the priming site for cDNA sequencing is highlighted. The vertical arrow marks the boundary of the *psbF* twintron and the 3′ *psbF* exon.

Intermediates in the splicing reaction, such as intron lariat-3′-exons could also be templates. In order to characterize specific intermediates in the twintron splicing pathway, the polymerase chain reaction (PCR) was used to amplify RNA processing intermediates, which were then cloned and sequenced. The location of the cDNA and PCR oligonucleotide primers is shown in Figure 1. The cDNA primer described above was used for first strand synthesis. To ensure that intermediates in the mRNA maturation pathway were being amplified, and not excised introns, a synthetic 20-mer from exon 3 of *psbE* was used for PCR-amplification. With this oligonucleotide, only precursors to the 1.2 kb *psbE-psbF-psbL-ORF42* mRNA containing either an intact twintron or a spliced intron F1 would be amplified.

Two double-stranded cDNA products were obtained (data not shown). The most abundant cDNA was 455 bp in length. This cDNA is the size expected for a pre-mRNA with a spliced F1 intron. The other cDNA was 1073 bp in length. This cDNA is as expected for a pre-mRNA with an unspliced twintron. The 455 bp cDNA was cloned as a phagemid designated pEZC1014, and sequenced. The DNA sequence data for a region of the plasmid DNA complementary to the pre-mRNA, including the twintron junction region, is shown in Figure 4. This sequence corresponds to the spliced domain V of intron F1 with the internal 618 nt intron F2 removed. It is the same sequence as the most prominent cDNA sequence ladder of Figure 3A. The remainder of the cDNA sequence was also determined (data not shown). It has the expected sequence for the 3′-end of *psbE*, a 9 nt intergenic spacer, *psbF* exon 1, and the spliced *psbF* intron F1.

The existence of an intron-within-an-intron, or twintron, and the location of the twintron junction sequence within domain V of intron F1 are confirmed by the cDNA cloning of the partially spliced pre-mRNA as the plasmid pEZC1014. In addition, the identification of a partially spliced, polycistronic pre-mRNA linked to the proximal *psbE* coding region, is evidence for a sequential *in vivo* splicing pathway. The partially spliced mRNA must arise from splicing of the *psbF* twintron in a precursor to the *psbE-psbF-psbL-ORF42* mRNA, and not from splicing of an excised twintron.

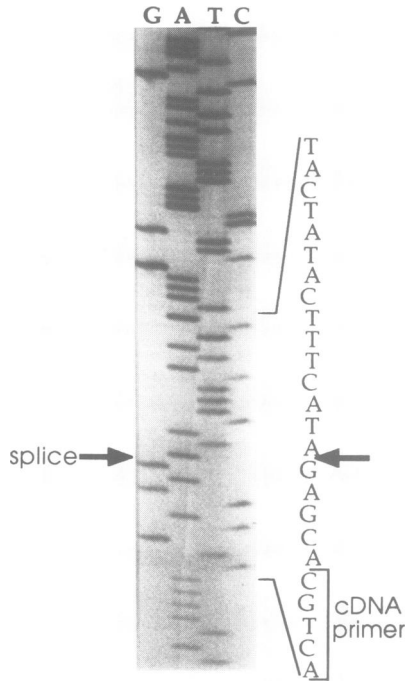

Fig. 4. The internal intron F2 is excised from the twintron of the *psbF* pre-mRNA. PCR was used to amplify precursors of the *psbF* mRNA. First strand cDNA synthesis was primed using the cDNA primer highlighted in Figure 3. PCR was used to amplify the resulting cDNAs using the cDNA primer and the *psbE* exon 3 primer (see Figure 1). An abundant cDNA of size 455 bp representing a pre-mRNA template with a spliced F1 intron (see Figure 1C) was cloned as the phagemid pEZC1014. The sequence data for this plasmid DNA encompassing the splice site is indicated. The DNA strand complementary to the pre-mRNA is shown. The splice site in domain V of intron F1 is indicated by arrows (see Figure 3B). A portion of the cDNA primer retained during the PCR cloning is shown in brackets.

Analysis of cytochrome b-559 pre-mRNAs and mRNAs

To determine more about how the twintron is processed *in vivo*, we used northern hybridization to detect intermediates in the splicing pathway. It had previously been determined that the *psaA-psaB-psbE-psbF-psbL-ORF42* genes are transcribed as a polycistronic operon, and that fully-spliced transcripts corresponding to hexacistronic, pentacistronic (*psaB-psbE-psbF-psbL-ORF42*), tetracistronic (*psbE-psbF-psbL-ORF42*), dicistronic (*psaA-psaB*), and monocistronic (*psaA, psaB*) mRNAs accumulate (Cushman, 1987; Manzara *et al.*, 1987; Christopher and Hallick, 1990). In order to detect precursors to mature *psbF* mRNA and excised introns, riboprobes specific to *psbE* intron 1 (E1), *psbE* intron 2 (E2), *psbF* external (F1) and internal (F2) introns, and exon-specific probes for *psbF* exon 2 and the *psbL-ORF42* locus were hybridized to membrane filter blots of chloroplast RNA. Total chloroplast RNA, and chloroplast RNA fractionated into 'soluble RNA' (sRNA) fraction and high molecular weight enriched RNA fractions (HMW RNA) were separated electrophoretically, transferred to nylon membranes, and hybridized individually with the 6 riboprobes. The results are shown in Figure 5.

The exon specific bands 1 (6.7 kb), 2 (3.4 kb), and 7 (1.0 kb), which were detected with both the *psbF* exon 2-specific and the *psbL-ORF42* locus probes, represent the fully spliced hexacistronic, pentacistronic (*psaB-psbE-psbF-psbL-ORF42*), and tetracistronic (*psbE-psbF-psbL-ORF42*) transcripts, respectively (Figure 5B) (Cushman *et al.*, 1988a; Christopher and Hallick, 1990). Note that further processing of the tetracistronic *psbE-psbF-psbL-ORF42* RNA is not observed. This is the end-product of the mRNA maturation pathway.

The intron-containing bands 3 (2.7 kb), 4 (2.3 kb), 5 (2.0 kb) and 6 (1.5 kb) were detected with the E1, E2, F1, and F2 intron-specific probes, as well as the *psbF* exon 2-specific and *psbL-ORF42* locus probes. We interpret these transcripts to be RNA processing intermediates of the 1.0 kb tetracistronic mRNA (band 1) with all four introns (band 3), three introns (band 4), two introns (band 5) and one intron (band 6) present. Since the splicing of *Euglena* chloroplast transcripts is not an ordered process (Koller *et al.*, 1985), the diffuse bands 4, 5, and 6 most likely represent populations of transcripts with different combinations of introns E1, E2, F1, and F2 present. Unspliced precursors of the hexacistronic *psaA-psaB-psbE-psbF-psbL-ORF42*

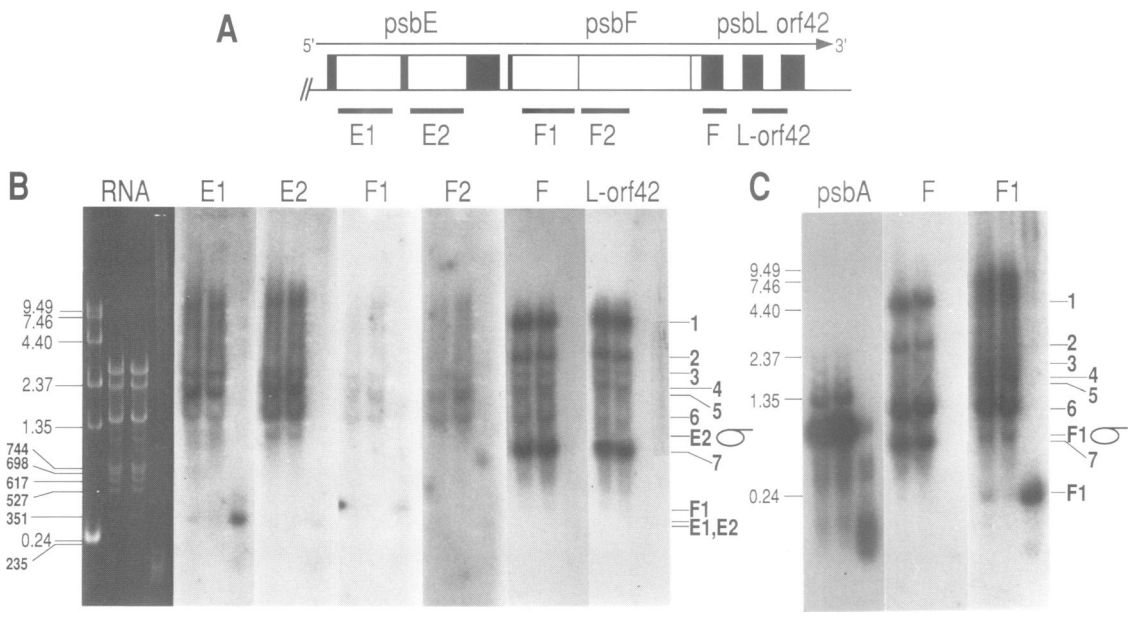

Fig.5. Analysis of *psbF* twintron-containing pre-mRNAs by northern hybridization. 3 μg of total chloroplast RNA, and chloroplast RNA that was fractionated into a 'soluble RNA' (sRNA) fraction and a high molecular weight enriched RNA fraction (HMW RNA) were electrophoresed on vertical 1.2% agarose gels containing 2.2 M formaldehyde. Fractionated RNAs were transferred to nylon membranes and hybridized to intron- and exon-specific probes. (**A**) The organization and structure of the *psbE-psbF-psbL-ORF42* genes and the location of the probes used in the northern hybridizations. Radiolabeled riboprobes complementary to the intron- and exon-specific sequences indicated at the bottom were prepared by T3 or T7 transcription from plasmids containing the corresponding sequences. (B) and (C) Northern analysis of *psbF* twintron-containing pre-mRNAs and control transcripts. The RNA panel shows the ethidium bromide stained gel of fractionated RNA markers (lane 1), total chloroplast RNA (lane 2), HMW RNA (lane 3), and sRNA (lane 4). The sizes of the RNA markers as well as internal chloroplast RNA markers (Schnare *et al.*, 1990) are indicated at the left. The panels corresponding to hybridization with the different riboprobes contained the same RNA lanes as the RNA panel without RNA markers. The hybridizations and post-hybridization washes in (**B**) were performed at a higher stringency than those in (**C**). The numbers at the right of (**B**) and (**C**) refer to *psbF* exon-specific transcripts (1,2, and 7) and intron-containing transcripts (3,4,5, and 6). Excised linear and/or lariat introns are also indicated.

RNAs are also observed, most notably with the E1 and E2 probes. Since unspliced precursors to both the hexa- and tetracistronic RNAs are present, the endonuclease cleavage event that produces the tetracistronic RNA is not ordered with respect to splicing.

The excised, linear introns were not readily detected with each probe in the total chloroplast or HMW-enriched RNA samples. However, RNAs corresponding to the linear introns E1 (0.36 kb), E2 (0.34 kb) and F1 (0.44 kb) were detected in the sRNA fraction enriched for small RNAs (Figure 5B and 5C). The intron F2 was not observed in this experiment. Its relatively low abundance may reflect either a low efficiency of conversion from lariat to linear form, or a more rapid degradation of this intron. To better distinguish between linear and lariat RNA molecules and to obtain more information on the twintron RNA splicing pathway, a novel, 2-dimensional gel system was employed.

Splicing of cytochrome b-559 pre-mRNA occurs via lariat intermediates

To determine if the splicing of chloroplast group II introns occurs via lariat intermediates, analogous to the fungal mitochondrial group II introns, a northern analysis of chloroplast RNA separated by electrophoresis in transverse linear gradient polyacrylamide gels was undertaken (see Materials and methods). The transverse linear gradient polyacrylamide gel system takes advantage of the fact that under denaturing conditions the electrophoretic mobility of lariat-containing RNA molecules changes much more dramatically as a function of polyacrylamide concentration than that of linear RNAs (Grabowski *et al.*, 1984). In 8% polyacrylamide−8 M urea, circular RNAs have slower mobility than like-sized linear molecules. By contrast, in 3.5% polyacrylamide−8 M urea, linear RNAs have a slower mobility than like-sized lariat molecules. In a 3.5−8% transverse gradient polyacrylamide gel, lariat molecules will cross over linear molecules at some point across the gradient. The point at which a lariat molecule crosses over a linear molecule depends on the size of the lariat. Total chloroplast RNA was electrophoresed through a gradient gel of 3.5−8%, transferred by electroblotting to nylon membranes, and hybridized with the 6 intron- and exon-specific riboprobes described above. The ethidium bromide stained gels (left-RNA panels), the northern blots (middle-Hybrids panels), and an optimized, computer-aided trace of the northern blots (right-Analysis panels) for each probe are shown in Figure 6. The RNAs detected via ethidium bromide staining are the 16S and 23S chloroplast rRNAs, the cytoplasmic 18S rRNA, and subfragments of the cytoplasmic 28S rRNA of known size (Schnare and Gray, 1990). They serve as internal standards for linear RNA size determination.

The intron-specific probes each detected a common linear RNA species of approximately 3.0 kb (band a in Analysis panels). This species is interpreted as the tetracistronic pre-*psbE-psbF-psbL-ORF42* mRNA with all four introns. A zone

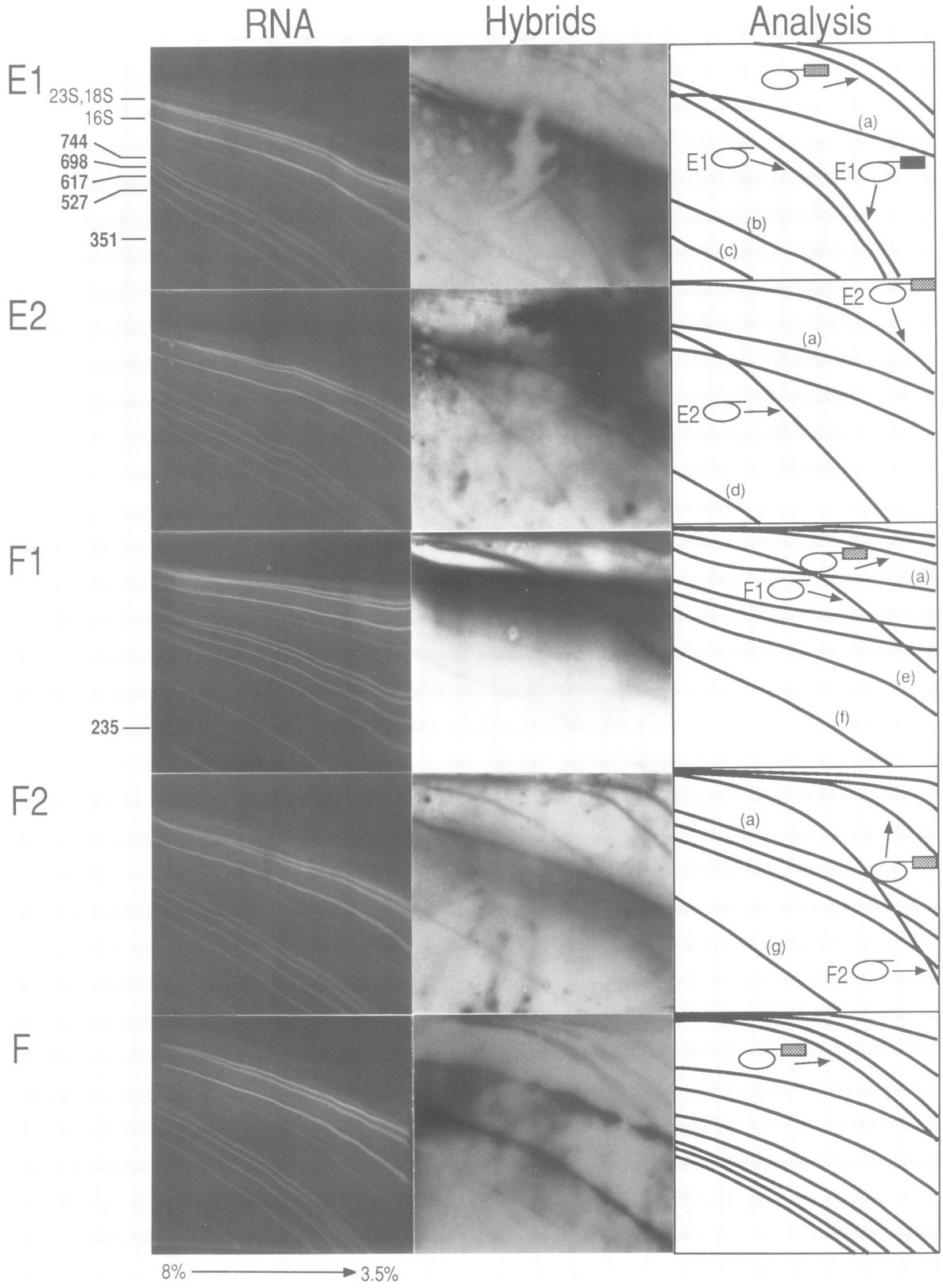

Fig.6. Splicing of the internal intron F2 and the external intron F1 occurs via lariat intermediates. 100 μg of total chloroplast RNA was fractionated on 3.5−8% transverse linear gradient polyacrylamide−8 M urea gels (see Materials and methods). Fractionated RNAs were electroblotted to nylon membranes and hybrized with the riboprobes described in Figure 5. The RNA panels show the ethidium bromide stained gels of fractionated chloroplast RNA. The internal chloroplast markers (Schnare *et al.*, 1990) described in Figure 5 are indicated at the left. The polyacrylamide gradient from left to right across the gel is 8%−3.5%. The hybrids panels show the northern hybridizations of the corresponding gels using the riboprobes indicated at the left. The analysis panels represent the major bands of the corresponding northern hybridization. The linear bands (a)−(g) are described in the text. The lariat RNA molecules for each analysis panel are indicated.

of smearing of multiple, low abundance linear RNA molecules beneath the tetracistronic pre-mRNA was also detected with each intron-specific probe. This region is interpreted as linear intermediates of the partially spliced tetracistronic pre-mRNA with various combinations of excised introns. Each of the intron-specific riboprobes also hybridized to RNA molecules that migrated at approximately the position of the linear forms of the excised introns (E1-band c, E2-band d, F1-band f, F2-band g). No linear form of an excised twintron was detected with either the F1- or F2-specific riboprobes.

High molecular weight, lariat-containing molecules (see upper right of Hybrids and Analysis panels) were observed with 5 different riboprobes. They are evident as curving bands that cross over the linear RNAs. With each of the intron specific probes (E1, E2, F1, F2), small lariat RNAs interpreted to be excised intron lariats, as well as larger lariat containing RNAs are observed. At least four different high molecular weight lariat molecules were also detected with the *psbF* exon-specific probe (panel F, Hybrids and Analysis). We interpret these larger molecules to be splicing intermediates that contain 5′-intron lariats with 3′-distal exons either spliced, partially spliced, or unspliced (the '2/3 molecule'). These lariat-containing RNA populations are heterogenous (panel F, Hybrids), as would be expected for non-ordered splicing of the 4 introns E1, E2, F1, and F2. Excised lariats were the most abundant of the lariat-containing RNA species, as well as the most abundant forms of the excised introns. The points at which these lariats crossed over a common linear molecule (i.e. the tetracistronic pre-mRNA, band a in Analysis panels) could be correlated with the expected size of the lariat molecules. The group II introns have putative branch sites for lariat formation at an A-residue in domain VI, 7 or 8 nt from the 3′-boundary of the intron. The largest of the introns (F2, 618 nt,) crossed the *psbF* pre-mRNA closest to the 3.5% acrylamide side of the gel (panel F2, Analysis). The next largest intron (F1, 424 nt) crossed over the pre-mRNA at a point near the middle of the gradient (panel F1, Analysis). The E1 lariats, the smallest of which is 350 nt, crossed the pre-mRNA very near to the 8% acrylamide side of the gel, while the smallest intron E2 (326 nt) did not cross over the *psbF* pre-mRNA (panels E1 and E2, Analysis). The E1-specific probe hybridized to two lariats that were very similar in size. We interpreted the largest lariat as the E1 lariat covalently linked to the fully-spliced downstream exons. The smaller lariat represented the excised E1 intron. The E2 probe detected only one lariat species interpreted to be the excised E2 intron. The F1- and F2-specific riboprobes only detected one lariat molecule that corresponded to their respective excised introns.

Discussion

psbF twintron is derived from two group II introns

The *psbF* gene of *Euglena gracilis* chloroplast DNA, which encodes the β-subunit of cytochrome b-559 of the photosystem II reaction center, was found to contain a novel genetic element consisting of one group II intron inserted into domain V of a second group II intron. This intron-within-an-intron was correctly predicted by Michel *et al.* (1989) based on secondary structure models for domains I−VI of both the internal and external intron. We have suggested the name 'twintron' to describe an-intron-within-an-intron.

Group II introns of plant and algal chloroplasts and fungal and plant mitochondria have a conserved core secondary structure (reviewed in Michel *et al.*, 1989), represented as a central wheel, with six radiating helical segments designated domains I−VI. Many *Euglena* chloroplast introns have the same core group II secondary structure (Michel and Dujon, 1983; Michel *et al.*, 1989), including both the psbF1 and psbF2 introns of the twintron (Figure 2). Nevertheless, *Euglena* chloroplast group II introns are considerably smaller than self-splicing fungal mitochondrial introns and are not self-splicing. Many lack one or more identifiable domains I−IV of the group II core secondary structure, and have abbreviated versions of the domains I−IV that are present. For example, the *psbF* introns have abbreviated versions of subdomains C and D and lack subdomains A and B of domain I (Figure 2). All *Euglena* group II introns have normal domains V and VI. Jacquier and Michel (1987) defined two sequences, designated EBS1 and EBS2 within domain I of group II introns that basepair with sequences designated IBS1 and IBS2 at the 3′-end of the 5′-exon. These potential EBS1-IBS1 and EBS2-IBS2 pairings are conceptually present within a psbF2-psbF1 interaction (Figure 2). The insertion of intron psbF2 into domain V of

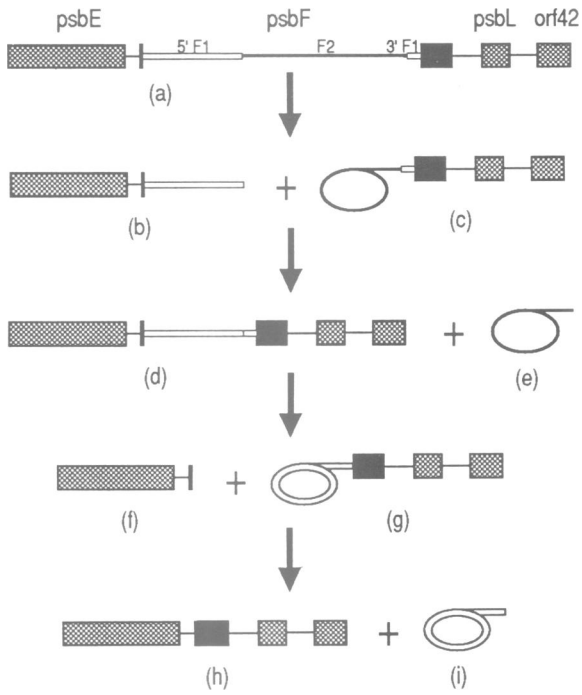

Fig. 7. A model for the sequential splicing of the *psbF* twintron. The *psbE-ORF42* pre-mRNA is shown. The filled boxes correspond to *psbF* exons. The hatched boxes depict the other genes in the pre-mRNA. The intercistronic regions are shown as thin lines. The 5′- and 3′-portions of the external intron F1 of the *psbF* twintron are represented by thin open boxes. The internal intron F2 is shown by a thick line. The contribution of the *psbE* introns to potential twintron-containing intermediates is ignored and represented as a hatched box. The internal intron F2 is first excised via a lariat-containing intermediate, which is then resolved into an RNA with spliced F1 intron in domain V and free F2 lariat. The external intron F1 is then excised via a lariat-containing intermediate. This intermediate is then resolved into an RNA that contains a fully spliced *psbF* gene and free F1 lariat. Not shown is the conversion of F1 and F2 lariats into linear intron forms. The RNA molecules (a)−(i) are described in the text.

intron psbF1 to form a twintron is significant because an intact domain V is essential to the normal splicing mechanism. Knowledge of the group II splicing mechanism is based largely on studies on the self-splicing introns of yeast mitochondrial genes (Peebles *et al.*, 1986; Perlman *et al.*, 1990; Van der Veen *et al.*, 1986). Splicing occurs by a pair of transesterification reactions. First, the 2′-hydroxyl of an adenosine 7 or 8 nt from the 3-splice boundary makes a nucleophilic attack on the 5′-nucleotide (usually guanosine) of the intron. The 5′-exon is released with a 3′-OH terminus, and the initiating A-nucleotide forms a 2′-5′-phosphodiester bond with the 5′-end of the intron, forming a lariat RNA. In the second reaction, the 3′-OH of the 5′-exon initiates a nucleophilic attack on the 3′-exon to produce the spliced exons and release the free intron lariat. Domain V has an essential functional involvement in the splicing reaction (Jarrell *et al.*, 1988), activating hydrolytic cleavage at the 5′-exon-intron junction and delivering domain VI containing the nucleophilic A-nucleotide to the catalytic site. The introduction of a 618 nt group II intron into domain V of intron psbFI would be expected to block splicing of this intron, and necessitate a sequential twintron splicing mechanism.

psbF twintron RNA splicing mechanism by sequential splicing of group II introns

To confirm the existence of twintrons, and to characterize events in the *psbF* mRNA processing pathway, a combination of cDNA cloning and sequencing, and RNA hybridizations with intron-specific probes was used. It is clear from our results that the internal intron F2 can be excised from the pre-mRNA prior to excision of the F1 intron. To the limits of detection of these experiments, we found no conclusive evidence for excision of the intact twintron. Domain V is absolutely required for the first reaction in the mechanism of group II intron splicing (Jarrell *et al.*, 1988). Therefore, we argue that the internal intron F2 has disrupted the functional domain V of the external intron F1 and two sequential splicing events are required for the maturation of *psbF* pre-mRNA. In the first step, the 618 nt internal intron is removed from the twintron, and the resulting segments of the external intron are spliced. In the second step, the spliced 424 nt external intron is removed and the *psbF* exons are spliced to yield the translatable mRNA.

Based on the characterization of intermediates of twintron splicing and the generally accepted mechanism for group II intron splicing, a model for the *psbF* twintron splicing mechanism is shown in Figure 7. The proposed RNA intermediates detected by cDNA and/or hybridization analysis are labeled (a)−(i). *psbF* mRNA maturation occurs within a polycistronic pre-mRNA, either tetra-, penta-, or hexacistronic. For simplicity, only *psbF* splicing events are considered. The initial *psbF* mRNA precursor contains the intact twintron (Figure 7a). Evidence for this precursor include its detection with intron F1 and F2 specific probes in northern hybridizations of 1D-gels (Figure 5), and transverse gradient gel electrophoresis (TGGE) (Figure 6). The twintron-containing pre-mRNA was also amplified as a cDNA via the polymerase chain reaction, and detected by cDNA sequence analysis (Figure 3). The initial splicing event is predicted to be cleavage at the 5′-boundary of the internal, or F2 intron, and release of the F2 intron as an intron lariat attached to the 3′-segment of the split F1-intron and the distal exon sequences (the '2/3 molecule') (Figure 7c). High molecular weight intron lariat species were detected by TGGE with an F2 intron-specific probe (Figure 6). However, it is not possible to determine which intron initiated the lariat in these molecules. The fact that the *psbF* exon-specific probe also detects high molecular weight lariats supports the existence of the 2/3 molecules. The branch point for lariat formation is inferred to be at the bulged A-residue in domain VI of intron F2 (Figure 3C) from the prominent stop in the cDNA sequence analysis (Figure 3A). The '1/3 molecule' linear splicing intermediate containing the 5′-segment of intron F1 and upstream exon sequences (Figure 7b) may also be evident in the TGGE analysis. Such an intermediate with a fully spliced *psbE* mRNA would have an expected length of 640 nt. A linear intermediate is detected with the F1-intron probe with mobility between that of the 698 and 617 nt internal RNA standards (Figure 6, band e panels F1-Hybrids and Analysis).

The second step in the sequential splicing pathway is ligation of the 5′- and 3′-segments of intron F1 to yield polycistronic pre-mRNA containing the partially-spliced twintron (Figure 7d), and release of the F2-intron as a lariat (Figure 7e). The identification of the partially-spliced twintron is the key experimental result confirming the existence of the twintron and a sequential splicing mechanism. It was identified by direct primer extension, cDNA sequence analysis with chloroplast RNA as template (Figure 3). The partially spliced twintron was also amplified by the polymerase chain reaction, cloned as a cDNA containing a portion of *psbE*, and sequenced, confirming that sequential twintron splicing occurs in the pre-mRNA maturation pathway.

The remaining steps in the twintron splicing pathway are comparable to the splicing of a normal group II intron, involving formation of the F1-intron lariat-3′-exon splicing intermediate, and subsequent ligation of the *psbF* exons with release of the F1-intron as an intron lariat (Figure 7,f−i). The F1-intron containing intermediates, and the fully-spliced mRNA were detected by northern hybridization and TGGE analysis. Correct splicing of the *psbF* exons was earlier confirmed by primer extension cDNA sequence analysis (Cushman *et al.*, 1988a).

Evolution of twintrons

How was the *psbF* twintron formed? One reasonable hypothesis is that the psbF2 intron was inserted into a progenitor *psbF* gene within a pre-existing psbF1 intron. This hypothesis is compatible with the idea that chloroplast genes have acquired introns in their evolutionary descent from an ancestral progenitor that lacked introns (Gingrich and Hallick, 1985). At least in the case of *Euglena*, intron addition seems to be involved in the evolution of chloroplast genes. Several group I introns have been shown to be 'infectious,' and this process is mediated by site-specific endonucleases encoded within the introns (reviewed in Perlman *et al.*, 1990). Some group II introns have been found to encode open reading frames with similarity to retroviral reverse transcriptase, including two yeast mitochondrial introns that have been shown genetically to be transmitted to other alleles lacking the introns (Meunier *et al.*, 1990). A plausible mechanism for group II intron transposition could involve a three step process of events

that are known to occur in organelles. The first step could involve a reverse splicing reaction mediated by the EBS1-IBS1 and EBS2-IBS2 interactions. This process has been shown to occur *in vitro* for group II introns (Augustin *et al.*, 1990). A cDNA copy of the product of reverse splicing could then be formed by reverse transcription. Finally, the cDNA could replace the pre-existing gene by homologous recombination, a process that occurs during chloroplast and mitochondrial transformation (Boynton *et al.*, 1988; Johnston *et al.*, 1988).

Euglena gracilis chloroplast DNA has many more introns than any other well-studied chloroplast genome. There are presently more than 50 known group II introns (R.B.Hallick, unpublished observation), and at least 50 group III introns (Christopher and Hallick, 1989; R.B.Hallick, unpublished observation). We estimate that at least 25–30% of the single copy region of *Euglena* chloroplast DNA encodes introns. The intron content of the *psaA* operon is even higher. The 13 introns of the 11.8 kbp *psaA-psaB-psbE-psbF-psbL-ORF42* operon total 6324 bp, or 54% of the DNA. If introns are being added to genes during chloroplast gene evolution, there is certainly a large target size for twintron formation within the *Euglena* chloroplast genome. The insertion of one intron into a pre-existing intron is necessary but not sufficient for twintron formation. The internal intron must also maintain its identity as a functional intron. One hypothesis to account for such an event is that upon insertion, the internal intron has disrupted an essential region of the external intron required for its excision. Therefore, a sequential intron splicing mechanism must occur in order to generate a translatable mRNA.

Do other types of twintrons exist? Given the large proportion of introns in the chloroplast genome of *Euglena*, it is likely that other twintrons also exist. The *psbF* twintron is a group II intron inserted into another group II intron. Based on an analysis of the boundary sequences of each intron, Christopher and Hallick (1989) proposed that the 409 nt *rps3* intron 1 was a group II intron inserted into a group III intron. This might be a mixed group II-group III twintron. We are in the process of testing this hypothesis.

If intron transposition is an active mechanism in the evolution of genes, one might expect remnants of twintrons in other kinds of pre-mRNAs. However, an internal intron would usually lose its identity when inserted into a non-essential region of an external intron, unless there was some advantage to regulate its splicing. Such a situation would maintain the original open reading frame while allowing genetic drift of the internal intron until an altered splicing event produces a transcript that becomes advantageous. This is the argument for exploitation of duplicated exons during the evolution of alternatively spliced genes (Smith *et al.*, 1989). In some cases, alternatively spliced pre-mRNAs with multiple 5′ and 3′ splice sites could have possibly been formed by intron insertion. A process of genetic drift and regulated splicing of the internal intron could have led to an alternatively spliced gene in which splice-site selection has major functional consequences to the expression of this gene.

Materials and methods

Molecular cloning

DNA sequence coordinates (see below) for subcloning have been described (Figure 2, Cushman *et al.*, 1988a). The recombinant plasmid pEZC704.6 was used as the source of DNA for cloning sequences specific for *psbE* introns 1 and 2. The 281 bp *psbE* intron 1-specific *Ssp*I–*Rsa*I fragment (positions 102–382), and the 264 bp *psbE* intron 2-specific *Dra*I fragment (positions 476–739) from pEZC704.6 were gel purified using GeneClean (BIO 101) and blunt-end ligated into the *Hinc*II site of Bluescript (−) (Vector Cloning Systems). The resulting plasmids were designated pEZC704.61 and pEZC704.62, respectively. The *psbF* exon-specific sequences (positions 2022–2128) were cloned by releasing the 250 bp *Eco*RV fragment from pEZC704.10 and religating the resulting molecule. This plasmid was designated pEZC704.13. The psbF1-specific sequences corresponding to positions 1057–1324 were cloned by isolating a 328 bp *Hin*dIII fragment from pEZC704 deletion clone Δ158 (Cushman *et al.*, 1988a) and ligating this fragment into the *Hin*dIII site of Bluescript (−). The resulting plasmid was designated pEZC704.11. The psbF2-specific sequences corresponding to positions 1350–1582 were cloned by digesting pEZC704.10 with *Hin*fI and *Acc*I, treating the resulting DNA fragments with 50 units/ml S1 nuclease, gel purifying the 233 bp fragment using GeneClean, and blunt end ligating this fragment into the *Hinc*II site of Bluescript (−). The resulting plasmid was designated pEZC704.12. The plasmid pEZC270, which contains sequences specific to *psbL* and *ORF42* at positions 2301–2467, has been described previously (Christopher and Hallick, 1989).

RNA purification

RNA was isolated from purified chloroplasts as described by Christopher and Hallick (1989). High molecular weight RNA and 'soluble RNA' (sRNA) were prepared by isopropanol fractionation as described by Orozco (1982).

Primer extension cDNA sequence analysis

A purified oligo-deoxynucleotide primer 5′-ATTATTTCCTATAAAAC-TGC-3′ complementary to the RNA-like strand in F1 of *psbF* (positions 1969–1988) was synthesized at the University of Arizona Biotechnology Center. The primer was 5′ end labeled using polynucleotide kinase. A total of 1×10^7 d.p.m. of [^{32}P]primer was co-precipitated with 10 μg of chloroplast RNA and resuspended in 12 μl of 200 mM KCl, 10 mM Tris–HCl (pH 8.3 at 42°C). The sample was heated at 85°C for 3 min, then quickly cooled on ice. The mixture was then heated at 50°C for 5 min, transferred to 42°C for 1.5 h, and then slowly cooled to room temperature. The primer extension reactions were done essentially as described by Christopher and Hallick (1989). The reactions contained 575 μM of dATP, 770 μM of dTTP, 290 μM each of dCTP and dGTP, 333 μM of ddATP, 500 μM of ddTTP, and 166 μM each of ddCTP and ddGTP.

cDNA synthesis, PCR amplification and cDNA cloning

The purified oligo-deoxynucleotide primer 5′-ATTATTTCCTATAAAA-CTGC-3′ complementary to the RNA-like strand in F1 of *psbF* used above in the primer extension cDNA sequence analysis was used to prime first strand cDNA synthesis using a cDNA synthesis kit (BRL). The reaction contained 10 μg of DNA-free total chloroplast RNA and 200 ng of oligo-deoxynucleotide primer. A purified oligo-deoxynucleotide primer 5′-CAATGCTTTAGAACAAATGG-3′ complementary to the template-like strand in *psbE* exon 3 (positions 916–935) was synthesized at the University of Arizona Biotechnology Center. This primer was used to amplify the resulting cDNAs by polymerase chain reaction (PCR) using the Taq polymerase (Perkin-Elmer Cetus). The reaction contained the cDNA synthesis product from 5 μg of chloroplast RNA and 200 ng of each oligo-nucleotide primer. PCR amplification consisted of 1 min of denaturation at 94°C, 2 min of annealing at 40°C, and 3 min of elongation by Taq polymerase at 72°C for a total of 35 cycles. The 455 bp *psbF* PCR product was gel purified using GeneClean, treated with 50 units/ml S1 nuclease to blunt the ends, and digested with *Mbo*I, which cuts at position 935 in *psbE* exon 3. The PCR fragment was then cloned into the *Bam*HI and *Hinc*II sites of Bluescribe (+). The resulting plasmid was designated pEZC1014.

Northern analysis

3 μg of total chloroplast RNA, high molecular weight enriched chloroplast RNA, and the sRNA fraction were electrophoresed on vertical 1.2% agarose gels containing 2.2 M formaldehyde. Fractionated RNAs were transferred to Genescreen membranes (NEN, DuPont Co.). The membranes were prehybridized at 50°C for 18 h in 50% formamide, 5× SSPE, 5× Denhardt's solution, 1% SDS, and 100 μg/ml herring testes DNA. The membranes were probed with uniformly ^{32}P-labeled complementary RNAs synthesized *in vitro* from linearized plasmids containing gene-specific sequences using either T3 or T7 RNA polymerase. The hybridizations were carried out in 50% formamide, 5× SSPE, 5× Denhardt's, 1% SDS and 100 μg/ml herring testes DNA at 50°C for 18 h with 1×10^7 d.p.m. ^{32}P-labeled probe. The membranes were washed in 2× SSPE, 0.5% SDS at 55°C for 1 h and 0.1× SSPE, 0.5% SDS at 55°C for 1 h. Autoradiography

was done for 2–24 h at −80°C with intensifying screens on Kodak X-Xomat-AR X-ray film.

Transverse linear gradient PAGE

Total chloroplast RNA was fractionated on 3.5–8% transverse linear gradient polyacrylamide–8 M urea gels using a modification of the Margolis and Kenrick (1968) procedure described for proteins. The gel plates were assembled and turned 90° such that the well corresponded to the right side. The gel was then poured using a linear gradient maker with 3.5% and 8% acrylamide–8 M urea in each chamber. Following polymerization, the gel was returned to its original position such that the well was now on top, resulting in a linear polyacrylamide gradient from left to right across the gel. 100 μg of chloroplast RNA was then loaded across the entire top of the gel and subjected to electrophoresis. Fractionated RNAs were electroblotted to Genescreen membranes and hybridized using intron-specific riboprobes as described above.

Acknowledgements

We wish to thank our colleague Gunter Wildner for the suggestion of 'twintron' to describe an intron-within-an-intron. We also thank Carol Dieckmann and Roy Parker for critical reading of the manuscript. We thank our laboratory colleagues for many stimulating and helpful discussions, and Noreen Sleator for help preparing the figures. This work was supported by the National Institutes of Health.

References

Augustin,S., Müller,M.W. and Schweyen,R.J. (1990) *Nature*,**343**, 383–386.
Bass,B.L. and Cech,T.R. (1986) *Biochem.*,**25**, 4473–4477.
Boynton,J.E., Gillham,N.W., Harris,E.H., Hosler,J.P., Johnson,A.M., Jones,A.R., Randolph-Anderson,B.L., Robertson,D., Klein,T.M., Shark,K.B. and Sanford,J.C. (1988) *Science*, **240**, 1534–1538.
Cantrell,A. and Bryant,D.A. (1987) In *Progress in Photosynthetic Research IV*, **10**, 659–662.
Carrillo,N., Seyer,P., Tyagi,A. and Herrmann,R.G. (1986) *Curr. Genet.*, **10**, 619–624.
Cech,T.R. (1986) *Cell*, **44**, 207–210.
Cech,T.R. (1988) *Gene*,**73**, 259–271.
Chakhmakhcheva,O.G., Andreeva,A.V., Buryakova,A.A., Reverdatto,S.V. and Efimov,V.A. (1989) *Nucleic Acids Res.*, **17**, 2858.
Christopher,D.A. and Hallick,R.B. (1989) *Nucleic Acids Res.*, **17**, 7591–7608.
Christopher,D.A. and Hallick,R.B. (1990) *The Plant Cell*, **2**, 659–671.
Cushman,J.C. (1987) Ph.D. thesis, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA
Cushman,J.C., Christopher,D.A., Little,M.C., Hallick,R.B. and Price,C.A. (1988a) *Curr. Genet.*, **13**, 173–180.
Cushman,J.C., Hallick,R.B. and Price,C.A. (1988b) *Curr. Genet.*, **13**, 159–171.
Dekker,J.P., Bowlby,N.R. and Yocum,C.F. (1989) *FEBS Lett.*, **254**, 150–154.
Gingrich,J.C. and Hallick,R.B. (1985) *J. Biol. Chem.*, **260**, 16156–16161.
Grabowski,P.J., Padgett,R.A. and Sharp,P.A. (1984) *Cell*, **37**, 415–427.
Haley,J. and Bogorad,L. (1990) *The Plant Cell*, **2**, 323–333.
Hallick,R.B. (1989) *Plant Mol. Biol. Reporter*, **7**, 266–275.
Herrmann,R.G., Alt,J., Schiller,B., Widger,W.R. and Cramer,W.A. (1984) *FEBS Lett.*, **176**, 239–244.
Hird,S.M., Willey,D.L., Dyer,T.A. and Gray,J.C. (1986) *Mol. Gen. Genet.*, **203**, 95–100.
Ikeuchi,M., Koike,H. and Inoue,Y. (1989) *FEBS Lett.*, **253**, 178–182.
Jacquier,A. and Michel,F. (1987) *Cell*, **50**, 17–29.
Jarrell,K.A., Dietrich,R.C. and Perlman,P.S. (1988) *Mol. Cell. Biol.*, **8**, 2361–2366.
Johnston,S.A., Anziano,P.Q., Shark,K., Sanford,J.C. and Butow,R.A. (1988) *Science*, **240**, 1538–1541.
Keller,M. and Michel,F. (1985) *FEBS Lett.*, **179**, 69–73.
Koller,B., Clarke,J. and Delius,H. (1985) *EMBO J.*, **4**, 2445–2450.
Kolosov,V.L., Klezovich,O.N., Abdulaev,N.G. and Zolotarev,A.S. (1989) *Bioorg. Khim.*, **15**, 1284–1286.
Manzara,T., Hu,J.-X., Price,C.A. and Hallick,R.B. (1987) *Plant Mol. Biol.*, **8**, 327–336.
Margolis,J. and Kenrick,K.G. (1968) *Analyt. Biochem.*, **25**, 347–362.
Meunier,B., Tian,G.-L., Macadre,C., Slonimski,P.P. and Lazowska,J. (1990) In Quagliariello,E., Papa,S., Palmieri,F. and Saccone,C. (eds), *Structure, Function, and Biogenesis of Energy Transfer Systems*. Elsevier Biomedical Press, Amsterdam, pp. 169–174.
Michel,F., Jacquier,A. and Dujon,B. (1982) *Biochimie*, **64**, 867–881.
Michel,F. and Dujon,B. (1983) *EMBO J.*, **2**, 33–38.
Michel,F., Umesono,K. and Ozeki,H. (1989) *Gene*, **82**, 5–30.
Nanba,O. and Satoh,K. (1987) *Proc. Natl. Acad. Sci. USA.*, **84**, 109–112.
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.
Orozco,E.M. (1982) In Edelman,M., Hallick,R.B., and Chua,N.-H. (eds), *Methods in Chloroplast Molecular Biology*. Elsevier Biomedical Press, Amsterdam, pp. 419–433.
Pakrasi,H.B., Williams,J.G. and Arntzen,C.J. (1988) *EMBO J.*, **7**, 325–332.
Peebles,C., Perlman,P., Mecklenburg,K., Petrillo,M., Tabor,J., Jarrell,K. and Cheng,H.-L. (1986) *Cell*, **44**, 213–223.
Perlman,P.S., Peebles,C.L. and Daniels,C. (1990) In Stone,E.M. and Schwartz,R.J. (eds), *Intervening Sequences in Evolution and Development*. Oxford University Press, New York, 112–161.
Schnare,M.N. and Gray,M.W. (1990) *J. Mol. Biol.*, **215**, 73–83.
Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwongse,J., Obokata,J., 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.
Smith,C.W.J., Patton J.G., and Nadal-Ginard,B. (1989) *Annu. Rev. Genet.*, **23**, 527–577.
Tae,G.S., Black,M.T., Cramer,W.A., Vallon,O. and Bogorad,L. (1988) *Biochemistry*, **27**, 9075–9080.
Tae,G.S. and Cramer,W.A. (1989) *FEBS Lett.*, **259**, 161–164.
Van der Veen,R., Arnberg,A., Van der Horst,G., Bonen,L., Tabak,H. and Grivell,L. (1986) *Cell*, **44**, 225–234.
Webber,A.N., Packman,L., Chapman,D.J., Barber,J. and Gray,J.C. (1989) *FEBS Lett.*, **242**, 259–262.
Westhoff,P., Alt J., Widger,W.R., Cramer,W.A. and Herrmann,R.G. (1985) *Plant Mol. Biol.*, **4**, 103–110.
Widger,W.R., Cramer,W.A., Hermodson,M. and Herrmann,R.G. (1985) *FEBS Lett.*, **191**, 186–190.

Received on October 8, 1990; revised on November 19, 1990