A group III twintron encoding a maturase-like gene excises through lariat intermediates

Donald W. Copertino1,2+, Edina T. Hall2§, Fredrick W. Van Hook2, Kristin P. Jenkins2 and Richard B. Hallick1,2*

1Departments of Biochemistry and 2Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

Received December 9, 1993; Revised and Accepted February 10, 1994

ABSTRACT

The 1605 bp intron 4 of the Euglena gracilis chloroplast psbC gene was characterized as a group III twintron composed of an internal 1503 nt group III intron with an open reading frame of 1374 nt (ycf13, 458 amino acids), and an external group III intron of 102 nt. Twintron excision proceeds by a sequential splicing pathway. The splicing of the internal and external group III introns occurs via lariat intermediates. Branch sites were mapped by primer extension RNA sequencing. The unpaired adenosines in domains VI of the internal and external introns are covalently linked to the 5' nucleotide of the intron via 2'-5' phosphodiester bonds. This bond is susceptible to hydrolysis by the debranching activity of the HeLa nuclear S100 fraction. The internal intron and presumptive ycf13 mRNA accumulates primarily as a linear RNA, although a lariat precursor can also be detected. The ycf13 gene encodes a maturase-like protein that may be involved in group III intron metabolism.

INTRODUCTION

The complete DNA sequence of the Euglena gracilis chloroplast genome is known (1). The chloroplast genes of Euglena contain 155 introns, including 74 group II introns. Many of the Euglena group II introns are smaller than those found in other organisms because domains I–IV are abbreviated (2–4). All possess a distinct catalytic domain V, and domain VI, which includes the unpaired adenosine residue that initiates nucleophilic attack at the 5' splice junction. Euglena chloroplast genes also contain 64 introns of a unique class designated group HI (5). Group HI introns are abbreviated group II introns that have retained group II-like 5' and 3' boundary sequences and a domain VI-like structure, but lack domains II–IV (2, 6, 7). It is not known if the putative group HI domain VI has a functional role during splicing in lariat formation. In many group HI introns, the region between the 5' boundary and domain VI resembles group II intron domain ID. Group III introns also occur in the plastid genes of the closely related nonphotosynthetic euglenoid Astasia longa (8, 9).

Euglena chloroplast genes also contain 15 twintrons, introns-within-introns, that are sequentially spliced (3). There are twintrons with one intron internal to another intron (2, 6, 10), and complex twintrons with multiple internal introns (11; for review, see 3). Group II and group III introns can occur as internal and/or external introns of twintrons. Excision of internal introns occurs prior to excision of external introns. The excision of internal group HI introns of twintrons can proceed from multiple 5' and 3' splice sites (6, 11). Internal introns are believed to interrupt functional domains of external introns, necessitating sequential splicing (3, 4).

The psbC gene of the Euglena gracilis chloroplast DNA, which encodes a 44 kDa protein component of the photosystem II reaction center (12, 13), contains an intron with an open reading frame (ORF) of 1374 nt (ycf13, 458 codons). It had previously been proposed that this intron is a typical group HI intron (14). We examined the splicing of the 1.6 kb intron by direct RNA sequencing, cDNA cloning and sequencing, and northern hybridization. This intron is a group HI twintron, and ycf13 is encoded within the internal group HI intron. Splicing of the internal and external group HI introns occurs via lariat intermediates, establishing a functional role for group HI domain VI.

MATERIALS AND METHODS

cDNA cloning and sequencing

Primers were synthesized by either The University of Arizona Biotechnology Center or The Midland Certified Reagent Company. Chloroplast RNA (ctRNA) was isolated and isopropanol-fractionated from purified chloroplasts as described (10). DNA sequence coordinates (see below) have been described (1; EMBL accession number X70810, release 37, version 18). Two synthetic deoxynucleotide primers complementary to the RNA-like strand (cDNA primers) at the psbC intron 4-exon 5 boundary (5'-GGGGATCCGCTACATGAGCTTAATTTAG-3', 10666 North Torrey Pines Road, La Jolla, CA 92037 and 1Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA)

*To whom correspondence should be addressed

Present addresses: +Department of Neurobiology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037 and §Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
positions 20234–20253), and in psbC exon 5 (5'-GGAA-CAAAATGAGCTACTTCTC-3', positions 20300–20319) were used to prime cDNA synthesis with total ctRNA as template (2). Another deoxyoligonucleotide 5'-GGAATCTATGACCTTAT-AAAAATGGATCC-3' of the RNA-like strand (PCR primer) in psbC exon 4 (positions 18596–18624) was used to amplify the resulting cDNAs by the polymerase chain reaction (PCR) (Perkin-Elmer, Cetus) as described (10). The PCR product derived from the partially spliced psbC RNA was digested with Xba I and BamHI, gel-purified by crush and soak (15), and cloned into the Xba I and BamHI sites of plBluescript KS- vector (Stratagene Cloning Systems). This plasmid DNA was designated pEZC1041 (Figure 1). The PCR product derived from the fully spliced psbC RNA was directly cloned into ddT-tailed plBluescript KS- (16). This plasmid DNA was designated pEZC1091 (Figure 1). The plasmid DNA pEZC1041.1 was generated by digesting pEZC1041 with HincII and SmaI, releasing a 40 bp fragment of the multiple cloning site that contains the EcoRI site, and religating the remaining plasmid DNA. To synthesize an external group III intron-specific riboprobe, pEZC1041.1 was linearized at the EcoRI site within the 5' region of the external intron (position 18648). This plasmid DNA also contains 10 nt of the 3'-exon as part of the cDNA primer. The plasmid cDNA inserts were completely sequenced on both strands by dideoxy DNA sequencing using the Sequenase kit (U. S. Biochemical).

Genomic cloning and sequencing

*Euglena gracilis* chloroplast DNA (ctDNA) was isolated as described (17). The cDNA primer at the psbC intron 4-exon 5 boundary and the PCR primer in exon 4 described above were used to amplify the corresponding region of the chloroplast genomic DNA. 200 ng of ctDNA in 2 μl of ddH₂O was boiled for 10 minutes, placed on ice, and amplified by PCR. 430 ng of cDNA primer and 400 ng of PCR primer were included in the amplification. Prior to amplification, a 2 minute 'hot start' at 80°C was done in the absence of Taq polymerase. The Taq polymerase was then added and amplification proceeded for 30 cycles. The DNA was denatured at 94°C for 1 minute, annealed at 42°C for 1 minute, and extended at 72°C for 2 minutes. The PCR product was then digested with *Xba* I and *Bam*HI, gel-purified by GENECLEAN (BIO 101), and cloned into the *Xba* I and *Bam*HI sites of plBluescript KS-. The plasmid DNA was designated pEZC2000. The internal group III intron-specific sequences corresponding to positions 19154–19587 were cloned by gel-purifying the 437 bp *Hind*III/EcoRI fragment from pEZC2000 using GENECLEAN, and ligating the fragment into the *Hind*III and *Eco*RI sites of the plBluescript KS-. The resulting plasmid DNA was designated pEZC2000.1. Identity of plasmid DNA inserts were confirmed by dideoxy DNA sequencing from the ends of the inserts, and by restriction endonuclease mapping.

Northern blot analysis

2 μg each of total ctRNA, high molecular weight-enriched ctRNA (HMW ctRNA), and the isopropanol-soluble (sRNA) fraction were subjected to electrophoresis through a 1.2% agarose gel containing 1×MOPS and 2.2 M formaldehyde using the optimized surface tension procedure (18). Fractionated RNAs were transferred to Genescreen (NEN, DuPont Co.) membranes and hybridized with riboprobes corresponding to intron-specific plasmid inserts generated *in vitro* by T3 or T7 RNA polymerase (see cloning above and Figure 2). The prehybridization, hybridization, and post-hybridization procedures were performed as described (10).

RNase H digestion and debranching of RNA lariats

12 μg of HMW ctRNA was dried with or without 500 ng of a purified deoxyoligonucleotide primer 5'-CCTCGCGGTATAAAGTAAAGAGAG-3¢ complementary to the RNA-like strand in psbC intron 4 at positions 19555–19577. Pellets were resuspended in 10 μl 25 mM Tris–HCl, pH 7.5; 1 mM EDTA; 50 mM NaCl; heated at 68°C for 10 minutes, and slow cooled to 30°C. An equal volume of 2 μl RNase H buffer (40 mM Tris–HCl, pH 7.5; 20 mM MgCl₂; 100 mM NaCl; 2 mM DTT; 60 μg/ml BSA) with or without 0.5 units RNase H (Promega) was then added and the reactions were incubated at 30°C for 1 hour. Reactions were terminated by adding 100 μl stop mix (38.5 μg/ml tRNA, 20 mM EDTA, 300 mM NaOAc). The mixture was phenol extracted, and products were ethanol precipitated. Control RNase H experiments were performed using 2.2 × 10⁵ dpm of transcript synthesized *in vitro* using T7 RNA polymerase from pEZC2000 linearized with *Bam*HI. This transcript is 1688 nt in length (41 nt vector, 31 nt exon 3, 21 nt 5' portion of external intron, 1503 nt internal intron, 81 nt 3' portion of external intron, and 11 nt exon 4). Transcription, purification, and self-splicing of the 887 nt yeast mitochondrial intron a5γ in 0.5 M (NH₄)₂SO₄ from pJD20 was done as described (19). 6 mg aliquots of HMW ctRNA and purified a5γ lariat were treated with HeLa S100 nuclear extract for 1 hour as described (20). Reactions were subjected to electrophoresis through 4% polyacrylamide gels containing 8 M urea in 44.5 mM Tris–borate, 44.5 mM boric acid, 1 mM EDTA (0.5× TBE), electroblotted to a Genescreen membrane, and either exposed to film (*in vitro* controls) or hybridized with a riboprobe specific for the internal ycf13-containing group III intron as described above (see Figure 2).

Primer extension cDNA sequence analysis

Purified deoxyoligonucleotide primers 5'-CACTAATTTAATTTAACAATC astronaut AACTACACTTAC-3', 5'-GGAAACAAATTGAGCTACT-TC-3', and 5'-GAAAATTTGAAAAATTTCTACATCAAAAAATC-3' complementary to the RNA-like strand in psbC at positions 20202–20225 (intron 4), 20300–20319 (exon 5), and 18726–18745 (ycf13), respectively, were 5' end-labelled using polynucleotide kinase (BRL). 32P-labelled primer (1×10⁷ dpm) was dried with 12 μg of total ctRNA or HMW ctRNA, or 2 μg sRNA and resuspended in 12 μl of 200 mM KCl; 10 mM Tris–HCl, pH 8.3 at 42°C. Two 6 μg aliquots of *Euglena* HMW ctRNA treated with HeLa S100 nuclear extract were also used as template. The hybridization and primer extension reactions were done as described (5). The reactions contained 770 μM dATP and dTTP, and 290 μM dCTP and dGTP, plus one of the following: 500 μM ddATP or ddTTP, or 167 μM ddCTP or dGTP. The primer extension reactions were subjected to electrophoresis on 6% polyacrylamide gels containing 8 M urea in 0.5× TBE.

RESULTS
cDNA cloning and sequencing

The chloroplast DNA (cDNA) of *Euglena gracilis* contains four intron-encoded open reading frames (1, 3). One of these, designated ycf13 with 458 codons, is located within the 1605 nt intron 4 of photosystem II gene *psbC* (Figure 1A). This intron
follows the 5'--GU--AG-3' rule typical for nuclear pre-mRNA introns (21), and was reported to contain structural features similar to group II introns (14). However, this hypothesis was made prior to the discovery of group III introns (22) and twintrons (10). In our view, the 3' structure is reminiscent of the domain VI-like region of group III introns (2, 6, 7). To test the hypothesis that the psbC intron 4 might be a new type of twintron, the polymerase chain reaction (PCR) was used with primers designed to amplify partially-spliced psbC pre-mRNAs. In control reactions, fully-spliced psbC mRNA, and genomic DNA representative of unprocessed pre-mRNA were also amplified. PCR products for partially-spliced and unprocessed pre-mRNAs are shown in Figure 1A.

The PCR product from partially spliced pre-mRNA was amplified, cloned (pEZC1041), and sequenced (Figure 1B). The cDNA insert of pEZC1041 is 143 bp in length and corresponds to a partially spliced RNA in which a 1503 nt internal intron has been excised, and a 102 nt intron, cDNA of fully spliced mRNA, containing portions of exons 4 and 5, was also cloned and sequenced (pEZC1091, Figure 1B). The exon 4-exon 5 splice boundaries are in agreement with an earlier report (14). As a control, a PCR product representing unprocessed pre-mRNA was cloned (pEZC2000), mapped by restriction endonuclease digestion (data not shown), and the boundaries of the insert were sequenced (Figure 1B).

The 1503 nt internal intron is also interpreted as a group III intron with an internal open reading frame of 1374 nt, the ycf13 gene, and 126 nt of non-coding sequence. Assuming ycf13 has 5' and/or 3' untranslated regions, the remaining non-coding region, of length approximately 100 nt, is the size expected for a group III intron (1, 5, 6). The internal intron also has a potential group III-like secondary structure (see below). Group III introns with internal ORFs, and ORFs within the internal intron of a twintron have not previously been described.

The identification of the partially spliced pre-mRNA is direct evidence that the 1605 nt psbC intron is a twintron, formed from the insertion of a group III intron into a group III intron. Identification of the partially spliced pre-mRNA is also consistent with a sequential in vivo splicing pathway (2). Since the RNA was amplified from the upstream exon, the isolated cDNA representing a partially spliced mRNA must be derived from splicing of a precursor mRNA, and not an excised twintron. No variation in splice sites was detected following splicing of either the internal or external group III introns (3/3 and 12/12 independent cDNAs, respectively). A consequence of the sequential RNA processing mechanism would be the accumulation of excised internal and external introns, rather than
twintron-sized products. To test this prediction, excised psbC introns were characterized by northern RNA hybridization.

**Northern blot analysis**

The *Euglena* chloroplast *psbC* gene is 11,403 bp in length, with 11 exons and 10 introns (1). Pre-mRNA transcripts of 2.1 and 2.5 kb, as well as a 1.5 kb species interpreted to be the fully spliced *psbC* mRNA were previously detected in the linear genomic plasmid DNA pEZC2000 (A) and *Euglena* *in vivo* HMW cRNA (B) were hybridized with an oligonucleotide specific to ycf13 and incubated with RNase H as indicated. A 32P-labelled purified lariat (and broken lariat) from the excised 887 nt yeast mitochondrial intron aI57 was included as a control for a linear and lariat RNA of known size. The *in vitro* RNase H controls and the purified aI57 lariat were directly subjected to autoradiography. The *in vivo* HMW cRNA samples were hybridized with the ycf13-specific riboprobe (see Figure 2). (C) A long exposure of the *in vivo* northern blot shown in (B).

The internal intron-specific probe hybridized predominantly to a 1.5 kb RNA (Figure 2A), the size expected for the excised internal ycf13-containing group III intron. This 1.4 kb RNA is distinct from the ycf13-specific 1.5 kb RNA. Intron-containing pre-mRNAs of 2.7 and 2.2 kb are also detected with the probe B. Pre-mRNAs of 5.8, 4.7, and 3.8 kb are also evident upon longer exposure of this blot (data not shown). The excised external group III intron was not detected.
in the three chloroplast RNA fractions with probe B. The riboprobe corresponding to the external group III intron which lacks the 5' exon 4 sequences but contains 10 nt from the 3' exon (probe C) hybridized with equal intensity to three RNA species (Figure 2C). The 1.4 kb RNA is distinct from the ycf13-specific RNA and is most likely the fully spliced psbC mRNA detected with probe B. The external group III intron probe also hybridized to intron-containing pre-mRNAs of 2.6 and 2.1 kb. The excised external group III intron was not detected in the three chloroplast RNA fractions with probe C.

Secondary structure model for the psbC twintron

Secondary structure models for the internal and external group III introns are shown in Figure 3. The most characteristic feature of group III introns is the group II intron-like domain VI at the 3'-end, with a bulged A -7 or -8 nt from the 3'-splice boundary that is the putative nucleophile for lariat formation. The insertion site of the internal intron is in the 5' region of the external intron, and may define another functional domain for group III introns. A secondary structure model for both introns of the psbC twintron was determined using the RNA FOLD program (23). Domain VI and the 3' end of the introns were excluded from the analysis (Figure 3). The internal group III intron is located within a predicted stem-loop structure in a U-rich bulge in the 5' region of the external group III intron. The structure is similar to that proposed for the external group III introns of the rps3 and the rpl16 twintrons (2, 3). A stem-loop structure in the 5' region of the intron is also predicted for the internal intron of the psbC twintron. The 1374 nt ORF is internal to this structure. An explanation for the ability of the unusually long 1503 nt internal intron to splice normally is apparent from the secondary structure model. The proposed intron core structure and helical domains might not be disrupted by the distal location of ycf13.

The 5' and 3' untranslated regions of ycf13 have not been defined. Although only the coding sequences of ycf13 were excluded from the computer analysis, the untranslated regions presumably would not contribute to the intron structure. Intron-encoded ORFs of yeast group II A introns are located in domain IV, which is not essential for intron excision in vitro (19, 24). Similarly, ycf13 must be positioned in a nonessential region of the internal group III intron. Both of the introns appear to lack exon-binding sites. The predicted folding of the internal intron of the psbC twintron also has a short helical domain between the 5' stem-loop and domain VI-like structures. This domain is absent from the external intron.

RNase H digestion of the psbC internal group III intron

Many intron-encoded open reading frames encode proteins known as maturases that function in splicing of their host intron. Some intron-encoded maturases are expressed as fusion proteins with the upstream exon, others are freestanding and are presumably translated independently of the upstream exon (for reviews, see 25–27). The group III twintron-encoded ycf13 is a freestanding ORF, and the excised intron containing the ORF accumulates to high levels (Figure 2). Group III introns of the rpl16, rps3, and rps18 twintrons accumulate as lariat or circular RNAs (6, 11, D.W.C. and R.B.H. unpublished). The internal ycf13-containing group III intron of the psbC twintron might also accumulate as a branched RNA or alternatively, as a linear, debranched molecule from which the ORF may be translated.

To test whether the ycf13-containing internal group III intron accumulates as a linear or lariat molecule, an internal intron-specific oligonucleotide primer complementary to the RNA-like strand within ycf13 was hybridized to Euglena HMW ctRNA and treated with RNase H. A circular or lariat intron would be cleaved via RNase H digestion of the primer/intron hybrid resulting in a linear molecule, whereas a linear species would be cleaved into two intron fragments. The reaction products were then analyzed by northern hybridization on acrylamide gels (Figure 4). Lariats migrate anomalously slowly compared to linear RNAs of equal size upon gel electrophoresis through high concentrations of acrylamide (28–30). A 1688 nt transcript of the psbC twintron (and portions of the surrounding exons, see Figure 1A) was synthesized in vitro and used as a linear RNA control. Purified lariat and linear (broken lariat) RNAs from the excised 887 nt yeast mitochondrial group II intron aI5y were also used as RNA mobility standards. The results of these experiments are shown in Figure 4.

The ycf13-specific probe (see Figure 2) hybridized to a transcript in the in vivo untreated and RNase H control samples which migrates as a linear (or debranched) 1.5 kb molecule, relative to the control samples of the 1688 nt psbC in vitro transcript and the linear and lariat 887 nt excised intron aI5y. RNase H digestion of the in vitro psbC transcript in the presence of the ycf13-specific oligo would result in two fragments of sizes 981 and 677 nt. These products are apparent in Figure 4A in the (+) oligo and (+) RNase H lane. In the (+) oligo and (+) RNase H in vivo samples, the ycf13-specific probe hybridizes to an RNase H cleavage product of 0.9 kb. The predicted size is 895 nt. This product is the 5' fragment of the internal intron, as expected for a linear internal intron substrate. The ycf13-specific probe does not contain sequences complementary to the 3' fragment of the internal intron.

A high molecular weight RNA that is slightly larger than the aI5y lariat is also detected with the ycf13-specific probe upon a longer exposure (Figure 4C). This RNA is a lariat or circular form of the internal ycf13-containing group III intron. Digesting ctRNA in the presence of the ycf13-specific oligonucleotide with RNase H results in linearization of this species. From these results we conclude that the internal ycf13-containing group III intron accumulates predominately as a linear molecule in steady-state ctRNA. However, a lariat or circular RNA, potentially the precursor of the linear species, is also present.

Primer extension RNA sequencing of the psbC twintron

Group II introns are excised as lariats, which form when the 2' hydroxyl of the unpaired adenosine within the domain VI helix is linked to the 5' nucleotide of the intron via a 2'-5' phosphodiester bond (28, 30). Group III introns are abbreviated versions of group II introns that contain a domain VI-like structure at the 3' end of the intron (2, 6, 7). The branch point for group III lariat formation has not been defined. To map the branch point, direct primer extension RNA sequencing was performed using a synthetic primer complementary to the RNA-like strand of the external intron. Total ctRNA, HMW, and the sRNA fractions were used as templates. The results are shown in Figure 5.

A ladder corresponding to a cDNA sequence of the unspliced twintron is obtained when total and HMW ctRNAs are used as template. A prominent termination in all lanes is also seen with these samples. However, this termination site is not present when the sRNA fraction or an in vitro RNA control is used as template (Figure 5). The termination site maps to the 5' nucleotide prior to the unpaired adenosine in domain VI of the internal group III intron, which is the predicted position for branch formation if
untreated mock HeLa
GATCNGATCNGATCN

Figure 6. The internal (A) and external (B) group III introns of the psbC twintron are excised via lariat intermediates. Untreated, mock-treated, and HMW cRNA treated with a HeLa nuclear S100 fraction were subjected to primer extension RNA sequencing as described in Figure 5. The arrows refer to the position of the external group III intron-internal group III intron, and exon 4-external group III intron boundaries. The branch sites hydrolyzed by the HeLa debranching activity are indicated by an arrow at the right.

it functions like domain VI of group II introns. The template for this cDNA is most likely the lariat intermediate for the internal intron. The 2'-5' phosphodiester bond between the unpaired adenosine within domain VI and the 5' splice site prevents primer extension by reverse transcriptase.

In order to confirm that the termination is due to branch formation, HMW cRNA was treated with the debranching activity of a HeLa nuclear S100 fraction (20), followed by primer extension analysis. As shown in Figure 6A, the termination site is susceptible to 2'-5' debranching activity. However, the 2'-5' phosphodiester bond is not completely hydrolyzed, persisting even after a second HeLa S100 treatment (data not shown). This result is probably due to partial inhibition of the enzyme by the RNA sample. Indeed, debranching of the yeast group II intron a15y as a positive control in the presence of Euglena HMW cRNA is also partially inhibited (data not shown). The 2'A-5'U phosphodiester bond is not as susceptible to cleavage by the HeLa debranching activity as the usual 2'A-5'G typically found in group II and nuclear pre-mRNA introns (31). The excision of the external group III intron was also examined by primer extension RNA sequencing using a primer complementary to the RNA-like strand of exon 5 (Figure 6B). One termination site is detected, which is absent following treatment with HeLa S100. This site maps to the expected position for lariat formation within domain VI during excision of the external group III intron. Interestingly, this 2'A-5'G bond is hydrolyzed more efficiently by the debranching activity than that of the internal group III intron. The difference in susceptibility of the HeLa debranching activity between the internal and external introns may be due to the size and sequence context of the lariat RNAs. To test whether processing other than debranching modifies ycf13 mRNA, the 5' end of the ycf13 mRNA was analyzed by primer extension RNA sequencing. The 5' end is the same as the 5'-end of the internal group III intron (data not shown). From these results we conclude that group III introns are excised via a lariat intermediate, and that group III intron domain VI is both the structural and functional equivalent to group II domain VI.

DISCUSSION

The excision of the 1605 nt intron 4 within the Euglena gracilis psbC gene was characterized by cDNA cloning and sequencing, northern hybridization, primer extension RNA sequencing, and debranching experiments. This intron is composed of an internal 1503 nt group III intron that contains an ORF of 458 codons (ycf13), and an external group III intron of 102 nt. Group III twintron excision was found to proceed by a two-step, sequential in vivo splicing pathway. A model for the RNA processing pathway is shown in Figure 7. Excision of both the internal and external group III introns occurs through a lariat intermediate. However, the ycf13-containing internal group III intron accumulates as a linear molecule. There is probably a precursor-product relationship between the excised internal intron-lariat and the ycf13 mRNA, proceeding via some uncharacterized debranching activity. Mohr et al (1993) have recently reported that domains of ycf13 are similar to the reverse transcriptase domains of group II intron ORFs, and a domain 'X' that may have an essential role in splicing. Thus, the identification of group III intron-encoded ycf13 of the psbC twintron extends the presence of maturase-like open reading frames to all classes of organelle introns.

Group III intron structure

The size range of a group III intron is 91-119 nt (5, 6). Since no group III introns outside of this size range were previously
known, the possibility existed that splicing was size-dependent. However, the 1503 nt internal group III intron of the psbC twintron is more than an order of magnitude larger than all other group III introns. Thus length alone is not the determinant of splicing. The most critical feature of group III introns may be the distance between the splice sites and the nucleophile in the core RNA structure. A similar relationship between size and core structure has been proposed for the two classes of yeast nuclear pre-mRNA introns. Most yeast nuclear pre-mRNA introns are between 250–1000 nt, whereas others are 100 nt or less. Intramolecular base-pairing between nucleotides near the 5' splice site and the branch site within the larger introns may result in structures comparable to the smaller introns (32). These interactions were shown to positively affect both splicing efficiency and splice site selection (33).

Group III introns also possess internal secondary structure between the 5' splice site and the branch site. The 5' regions of the internal intron of the rpl16 twintron and the external intron of the psbC twintron are almost identical to domain ID3 of group II introns (2, 3). The terminal loop of the 5'-domain of some group III introns may interact with the upstream exon in a tertiary interaction similar to the EBS1-IBS1 of group II introns (4). Other group III introns have potential stem-loops in the 5' region of the intron that lack EBS-like sequences. These 5'-domains may have evolved from domain ID and now function in a core structure required for splicing without a tertiary interaction with the exon. Several intron insertion sites among group III and complex twintrons are within these predicted 5'-helical domains (3). The internal intron of the psbC twintron also disrupts a potential stem-loop structure of the external intron (Figure 3). The internal introns may have been maintained because the structures they interrupt are essential for splicing of the external intron (3).

Group III introns excise as lariats

Based on secondary structure analysis, group III introns have been predicted to contain domain VI at their 3' ends. Six of ten twintrons involving external group III introns contain insertion sites within this domain, consistent with a functional role for this domain (3). Several group III introns have been shown to accumulate as lariat or circular RNAs (6, 11, D.W.C. and R.B.H. unpublished). Termination sites in the primer extension RNA sequencing of splicing intermediates of the psbC group III twintron map to the unpaired adenine in domain VI-like structure of the internal and external group III introns. Termination is sensitive to the debranching activity of a HeLa nuclear extract that hydrolyzes 2'-5' phosphodiester bonds (Figure 6). Therefore, group III introns are excised as lariats in which the unpaired adenine in domain VI is covalently attached to the 5' nucleotide of the intron in a 2'-5' phosphodiester bond. Furthermore, domain VI of group III introns is both structurally and functionally equivalent to domain VI of group II introns. An evolutionary and mechanistic relationship between group II and group III intron excision seems apparent.

The evolution of the group II and group III splicing systems in euglenoid plastids may parallel the evolution of nuclear pre-mRNA splicing (3). Group III introns represent abbreviated group II introns. Part or all of the group II intron catalytic core may be reconstituted for group III intron excision (3, 6). Many of the cis-elements required for group II intron excision, such as domain V, may be supplied as trans-acting RNAs for group III introns. Domain V interacts with domain VI in trans during in vitro group II intron excision, however, this interaction is very inefficient (34). In addition, a deletion analysis of substructures of the yeast group II intron al57 has demonstrated that domain V binds to domain I in trans and together with the 5' exon constitute the catalytic core of this group II intron (35). Domain V may interact with the 5' domain ID-like element of group III introns thereby activating the first reaction at the 5' splice site. The nucleophile within domain VI may arrive at the active site due to the relatively reduced size of the group III intron core structure. Nuclear pre-mRNA splicing requires trans-acting snRNPs which assemble on the intron to form the spliceosome (for reviews, see 36, 37). Important regions of some of the snRNAs of these ribonucleoproteins appear to be structural and/or functional cognates to the cis-encoded domains of group II introns (38–41). Nuclear and group III introns may be directly related to functional cognates to the cis-encoded domains of group II introns (38–41). Nuclear and group III introns may be directly related descendants of group II introns or may represent parallel evolution from group II introns (3). Other similarities between group III and nuclear introns include conserved 5' boundary sequences, lariat formation, lack of internal structure, and ability to use alternate splice boundaries.

Does ycfI3 encode a maturase?

Several group I and group II introns contain open reading frames that encode maturases (25–27, 42). These maturases are required for efficient splicing in vivo and are thought to bind and stabilize catalytically active RNA structures. The fungal mitochondrial group II intron-encoded maturases are highly conserved and contain domains with a high degree of similarity to reverse transcriptases (42–45). Reverse transcriptase activity associated with several of the yeast group II intron-encoded maturases that is involved in intron mobility has also been demonstrated (46). Many of the ORFs contain a conserved domain X and a Zn2+-finger-like region (42). The putative group II-intron encoded maturases within the plastid tmK genes from tobacco, pea, mustard, and rice contain only remnants of the blocks of amino acids that are diagnostic for reverse transcriptase-like sequences. These maturases lack the Zn2+-finger-like region, yet retain domain X. Consequently, domain X may be essential for the maturase activity of intron-encoded open reading frames (42).

Group III intron excision is also likely to require proteins in vivo. One potential group III maturase is ycfI3, which is encoded within the internal group III intron of the psbC twintron. Interestingly, the plastid DNA of the colorless, heterotrophic euglenoid Astasia longa contains an open reading frame (orf456) that has 55.5% identity and 84.8% overall similarity to the Euglena gracilis ycfI3 (8). Both of these ORFs encode domains with homology to the reverse transcriptase domain and domain X of other intron-encoded ORFs (42). However, the Astasia orf456 gene is not intron-encoded. The plastid DNA of Astasia is approximately 73 kb in size, and lacks most or all photosynthetic-related genes other than rbcl. The Astasia orf456 gene is on the same strand and in relatively the same location of the genome as the Euglena ycfI3 gene, given deletions of the psbA and psbC genes. Apparently, the Astasia plastid DNA has undergone deletions of dispensable genes, particularly those involved in the photosynthetic light reactions. Other highly conserved genes have been maintained, and therefore may be essential for plastid functions. The ycfI3 gene is also present within the psbC intron in four other species of Euglena (D.W.C. et al., unpublished). The plastid genes of Astasia longa and other euglenoid species also contain group III introns (8, 9, D.W.C.
et al. (unpublished). ycf13, which is group III twintron-encoded in Euglena, may be required for group III intron excision and/or mobility in Euglena and Astasia. A similar situation is found within the plastid genome of the nonphotosynthetic parasitic plant Epifagus virginiana. The matK ORF is closely related to the tmK introns (47). Presumably, matK has been retained since it may be required for the excision of other group II introns (47). We are currently testing whether ycf13 is necessary for group III intron excision and/or mobility.

ACKNOWLEDGEMENTS

We wish to thank Jenny Stevenson and Mike Thompson for critical reading of the manuscript. We also wish to thank Dr. Philip Perlman for the plasmid DNA pJD20 of the yeast mitochondrial intron a5y. Drs. Steven Mayer and Bernardo Nadel-Ginard for the HeLa S100 nuclear extract, and Giordano Caponigro and Denise Muhrad for the RNase H protocol. This work was supported by the US National Institutes of Health.

REFERENCES