A complex twintron is excised as four individual introns

Robert G. Drager¹ and Richard B. Hallick^{1,2,*}

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

Received February 16, 1993; Revised and Accepted April 14, 1993

EMBL accession no. Z11874

ABSTRACT

Twintrons are introns-within-introns excised by sequential splicing reactions. A new type of complex twintron comprised of four individual group III introns has been characterized. The external intron is interrupted by an internal intron containing two additional introns. This 434 nt complex twintron within a *Euglena gracilis* chloroplast ribosomal protein gene is excised by four sequential splicing reactions. Two of the splicing reactions utilize multiple 5'- and/or 3'-splice sites. These findings are evidence that introns with multiple active splice sites can be formed by the repeated insertion of introns into existing introns.

INTRODUCTION

The chloroplast genome of the photosynthetic protist *Euglena gracilis* contains at least 142 introns (1). This is seven times the 18–21 introns of land plant chloroplast genomes (2, 3, 4). Based on phylogenetic comparisons of prokaryotic and plastid genes, it has been proposed that the numerous *Euglena* introns, as well as most higher plant chloroplast introns, have arisen via intron insertion during evolution of the chloroplast genome (5, 6). The *Euglena* chloroplast introns are of two classes, group II introns also found in other chloroplasts and plant and fungal mitochondria, and group III introns which are unique to the plastids of *Euglena gracilis* and the non-photosynthetic protist *Astasia longa* (7, 8). The two classes of introns are readily distinguished by their characteristic features. *Euglena* group II introns are 277–671 nt in length and have highly conserved secondary structure (9, 10). Group III introns have a very narrow size range of 91–119 nt (7).

The *Euglena* chloroplast genome also contains introns-within-introns, termed twintrons. Three types of twintrons have been described: group II within group II, group II within group III and group III within group III (11, 12, 13). Twintrons are excised from pre-mRNAs by sequential splicing of the individual introns. The internal intron is excised first, thereby splicing the separate portions of the external intron. Subsequent excision of the external intron results in splicing of the exons. The existence of twintrons is consistent with the concept that group II and group III introns are mobile genetic elements. A three step process of reverse splicing, reverse transcription and homologous recombination is one possible mechanism of intron insertion and twintron

formation (11). If twintrons are formed by the insertion of a mobile intron into an existing intron, it is possible that intron insertion into twintrons could result in more complex introns. In principle, a complex intron could be formed from 3, 4 or more individual introns.

The sequence and gene structure of a *Euglena gracilis* chloroplast polycistronic operon encoding two chloroplast ribosomal proteins and four CF₀-CF₁ ATP synthase subunits has recently been determined (coordinates 30503–39956 EMBL accession #Z11874) (14). The six genes of this operon are interrupted by seventeen introns. Nine of these introns are 99–112 nucleotides in length, and were identified as group III introns. The other eight introns are 323 to 637 nt in length. Seven of the larger introns were determined to be group II introns based on their secondary structure (14). The 434 nt intron 2 of *rps18*, (Fig. 1, coordinates 39466–39899) is too large to be a group III intron, but does not have a group II secondary structure. We suspected that this intron might be a new type of twintron, comprised of four group III introns. This hypothesis was tested by PCR amplification and subsequent sequencing of partially spliced cDNAs and by northern hybridization analysis. This 434 nt intron in the ribosomal protein S18 mRNA precursor is excised as four distinct group III introns. The excised introns accumulate as circular or lariat RNA molecules. Two of these group III introns utilize multiple 5'- and/or 3'- splice sites. We use the term 'complex twintron' to describe an intervening sequence excised as three or more discrete introns. The existence of complex twintrons is evidence that larger introns can be assembled from smaller, individual introns, and is suggestive of a mechanism for the evolution of introns with multiple active splice sites.

MATERIALS AND METHODS

cDNA synthesis, amplification and cloning

Total RNA was isolated from frozen *Euglena gracilis* chloroplasts as previously described (14, 15). cDNAs were synthesized using specific oligonucleotide primers (University of Arizona Biotechnology Center) as described by Yepiz-Plascencia, et al. (16). cDNA primer #1 (5'GCGTGCTTGTTAACTTTTG3', complementary to coord. 39914–39895 EMBL accession #Z11874) spans the *rps18* intron 2-exon 3 boundary. cDNA primer #2 (5'GGAAGTAAGCCTAAATAGCGTGC3',

* To whom correspondence should be addressed at: Department of Biochemistry, University of Arizona, Biosciences West 524, Tucson, AZ 85721, USA

complementary to *coor.* 39931–39909) is within *rps18* exon 3 (see Fig. 2A). Resulting cDNAs were amplified by the polymerase chain reaction (PCR, Perkin-Elmer) with a primer in *rps18* exon 2 (5'-GAAGATTCATAAACTTTCAAGG3', *coor.* 39374–39395) as previously described (11, 12, 13, 14). As a control, purified *Euglena* chloroplast DNA (15) was also amplified using each cDNA primer with the *rps18* exon 2 PCR primer. All PCR reactions consisted of a 3 min. 80°C hot start followed by 25 cycles of 1 min. 94°C, 2 min. 50°C, 2 min. 72°C. PCR products were cloned into ddT tailed Bluescript KS(+) (17) by both shotgun cloning of total PCR products and by gel purification (GeneClean, Bio 101) of individual cDNAs. Cloned cDNAs were sequenced by standard methods using the Sequenase DNA sequencing kit (U.S. Biochemical).

Northern hybridization analysis

Preparation of the northern blots of low molecular weight *Euglena gracilis* RNA has been previously described (13). Briefly, 2 µg isopropanol soluble RNA (11) was fractionated by electrophoresis on 6% and 15% polyacrylamide slab gels containing 8M urea. Fractionated RNAs were transferred to Genescreen (NEN DuPont Co.) and hybridized to ³²P labelled RNA probes complementary to the entire *rps18* intron 2, *rps18* introns 2a and 2b or *Euglena* chloroplast 5S RNA.

RNA secondary structure determination

The most stable RNA structure of nucleotides 39562–39868, representing all of *rps18* intron 2b excluding the putative domain VI-like region, was determined using the FOLD program (18).

RESULTS

cDNA cloning and sequence analysis

The DNA sequence and structure of the *Euglena* chloroplast *rps18* gene is presented in Fig. 1. The positions of primers used for cDNA synthesis and PCR amplification of *rps18* intron 2 are shown in Fig. 2A. The products of PCR amplification of cDNAs are shown in Fig. 2B. Amplification of cDNAs synthesized from cDNA primer #1, which spans the intron–exon boundary and cannot amplify fully-spliced transcripts, results in PCR products of 560 bp, 450 bp, 340 bp, 230 bp and approximately 50 bp (Fig. 2B, lane 2). The four larger products are interpreted as the unspliced primary transcript and three partially spliced pre-mRNAs. The product of approximately 50 bp, as determined from cDNA sequence analysis, is an artefact of mis-priming during cDNA synthesis. Six nucleotides at the 3'-end of cDNA primer #1 (5'-CTTTTG-3') are complementary to nucleotides 39407–39412 (5'-CAAAAG-3') of *rps18* exon 2. Because this product only appears upon amplification of cDNAs and not upon amplification of genomic DNA with the same primers (Fig. 2B, lane 4), it is likely that the mis-annealing took place during cDNA synthesis at 37°C, rather than during the PCR reaction in which the annealing temperature was 50°C.

PCR amplification of cDNAs synthesized from cDNA primer #2, which is within *rps18* exon 3, results in four products of 560 bp, 450 bp, 340 bp and 230 bp that correspond to the cDNAs obtained with primer #1, plus an additional product at 120 bp (Fig. 2A, lane 3). The 120 bp band corresponds to the predicted cDNA product of the fully-spliced transcript. The sizes of the PCR amplification products of the *rps18* intron 2 cDNAs are consistent with the hypothesis that this 434 nt intron is excised as four individual group III introns, each 100–110 nt in length.

```

AAGATGTTTAACTTTATGAAATTTGTGATGTACTTTTAAATAAAATAAAATAAAATA
N K
TTATAAAATAAAATTTTAAATATAAACTTAAACAAAATTTTGAAGAAATTTTATC
39350
ATATACCCAGCCTTATTGATTATAAAATGTTAAATCTTGAAGATTCATAACTTT
(Y) P A L I D Y K N V N I L R R F I N F
39400
CAAGGTAAATAATTCCAAACGATTAAACAACTAAATTAACATATAACAGCACCT
Q G K I I P K R L N K P K L T Y K Q H R
39450
CTTCTAAGAAATCTTTTGTATTATTTAAATATTAAATAAAATTAATAATGACATAAT
L L R K S
39500
ATCAATAATCAACAAATAAAATTTTGTCTTCTTAAGAAAAACATATTATGTTT
39550
CACGTTAAATTTTAAATAATAGTAAATAATAGTGAATTAAACAAATAAAATGAG
39600
TATTCACAAAATTTTAAACAAATACATATTTTAAATAATCTTTATATTGAA
39650
AAACATTTAGTTATTTTAAATAATTTGTAAATGATTTCTTATTTTAAATAAT
39700
ACAAAATGTAAGATTAAATTTTAAATAATTAATTAATTAATTAATTAATTAAT
39750
AAATTAATAATTTTCTTGTAGATTAAATATACGCATTTTAAAGATTAAATTATG
39800
ACTACATTTTAAATAAGTATATCAAAAGTTAAACAGACCGCTATTAGGCTACTTC
V K Q A R Y L G L L P
39850
CTTTTAAACATAAAGATTTTAAATATCTTTAGTTTAA
F K T K D F F *

```

Figure 1. DNA sequence of the RNA-like strand of the *Euglena gracilis* chloroplast *rps18* gene, coordinates 39211–29972 EMBL accession # Z11874. Coding regions are designated by the single letter amino acid code directly below the first nucleotide of each codon. The codon of the tyrosine residue shown in parentheses is split by intron 1. Nucleotides of intron 1, a group III intron (*coor.* 39234–39334), and intron 2, the complex twintron (*coor.* 39466–39899), are shown in italics.

Amplification of *Euglena* chloroplast DNA with each cDNA primer and the exon 2 PCR primer results in only the 560 bp product (Fig. 2B, lanes 4 and 5).

The PCR amplified cDNAs were cloned and sequenced. The cDNA amplification products of 560 bp, 450 bp, 340 bp, 230 bp and 120 bp correspond to transcripts from which 0, 1, 2, 3 and 4 group III introns have been excised. These four introns are arranged in a complex pattern of two group III introns within the internal intron of a group III twintron (Fig. 2C). The external intron 2a is interrupted once, and therefore consists of two non-contiguous pieces. The second intron 2b is interrupted twice and consists of three non-contiguous pieces. The remaining two introns, 2c and 2d, are intact and internal to intron 2b. Based on the analysis of the cloned cDNAs, the RNA maturation pathway shown in Fig. 3 is proposed as the major route of excision of *rps18* intron 2. Partially spliced cDNAs from which intron 2c, or introns 2c and 2d, or introns 2b, 2c and 2d had been excised were characterized. All cDNAs from which only one intron had been excised lack intron 2c. These results are most consistent with the sequential *in vivo* excision of introns 2c, 2d, 2b and 2a, respectively. Excision of introns 2c and 2d results in the splicing of the three non-contiguous pieces of intron 2b into an intact intron. Subsequent excision of intron 2b results in the splicing of the two non-contiguous pieces of intron 2a into an intact intron, which is then excised to yield the fully-spliced *rps18* mRNA. However, from these data the excision of the entire 434 nt intron or intron 2b interrupted by 2c and/or 2d cannot be excluded, since some of the same cDNA products would result from these splicing events.

The sizes of individual introns 2a, 2b, 2c and 2d are 107, 109, 106 and 112 nt, respectively. Each of the four individual introns has the characteristic features of group III introns. These are size between 90–120 nt, 85–91% A+U base content, and a subset of conserved nucleotides characteristic of group II intron 5' splice sites. The 5' boundary sequence of each group III intron has a guanosine residue at position +5, and most also have a uridine at position +2. This is similar to the group II intron 5'-consensus of 5'-GUGYG. The 3'-boundary sequences are variable, but all

Figure 2. Analysis of the partially spliced *rps18* pre-mRNAs. (A) Positions of primers used for cDNA synthesis from total *Euglena gracilis* chloroplast RNA and PCR amplification of *rps18* intron 2 cDNAs. A PCR primer in *rps18* exon 2 was used to amplify cDNAs synthesized from cDNA primers #1 and #2. (B) Ethidium bromide stained 1.7% agarose gel of PCR products. Lane 1) ϕ X174 DNA digested with HaeIII. Lane 2) PCR amplification products of cDNAs synthesized from cDNA primer #1 and amplified with cDNA primer #1 and the exon 2 PCR primer. Lane 3) PCR amplification products of cDNAs synthesized from cDNA primer #2 and amplified with cDNA primer #2 and the exon 2 PCR primer. Lane 4) PCR products of *Euglena gracilis* chloroplast DNA amplified with cDNA primer #1 and the exon 2 PCR primer. Lane 5) PCR products of *Euglena gracilis* chloroplast DNA amplified with cDNA primer #2 and the exon 2 PCR primer. The sizes, in bp, of MW markers are shown on the left and the sizes of the PCR products are shown on the right. (C) Structure of the *rps18* complex twintron as determined by sequence analysis of partially spliced and fully-spliced cDNAs.

have an unpaired adenosine residue 7–10 nucleotides proximal to the 3'-splice site, within a potential structure similar to domain VI of group II introns (12, 13, 14).

Utilization of multiple 5'- and 3'-splice sites

Introns 2b and 2c are excised from multiple 5' and/or 3'-splice sites, however no variations were observed in the splice boundaries of introns 2a and 2d. Sequence data from cDNA clones from which both internal introns 2c and 2d have been removed are shown in Fig. 4. Intron 2c is excised from two different 5' splice sites 2 nt apart, and four different 3'-splice sites at four adjacent nucleotides. The excision of intron 2b occurs from two 5' splice sites 25 nt apart (data not shown). Both intron 2b 5'- sites are spliced to the same 3'- site. The frequency and splice boundaries of the observed cDNAs are presented in Table 1.

Internal introns are within potential secondary structures

Two potential secondary structures similar to features of group II introns have been described in group III introns. A sub-set of group III introns have structures resembling domain ID of group II introns (12) and all group III introns have a domain VI-like region near the 3' splice site (12, 14). The sites of insertion of introns 2b and 2d are within the domain VI-like regions of introns 2a and 2b, respectively (Fig. 5). Intron 2c interrupts intron 2b within a secondary structure determined by the FOLD

Figure 3. The proposed major pathway of *Euglena gracilis* chloroplast *rps18* RNA maturation as determined by sequence analysis of cloned cDNAs. Excision of intron 1 is independent of the multi-step splicing of intron 2, a complex twintron. Diagrams from top to bottom represent RNAs from which 0, 1, 2, 3 and 4 group III introns have been excised from *rps18* intron 2. Exons are shown as black boxes and introns as either white (1), 2a, 2c, 2d or gray 2b boxes. Coordinates of the most frequently observed individual group III intron splice sites are indicated by vertical arrows.

program (18), (Fig. 5B). This stem-loop structure is similar to domain ID of group II introns and contains a potential exon binding site (EBS) in the loop. Although intron 2c is excised from two 5'-splice sites and four 3'-splice sites, the nucleotide pairing interactions of the main stem of this structure in intron 2b remain unchanged.

Northern hybridization analysis

The size and topology of the group III introns excised from the *rps18* complex twintron were determined by northern blot analysis. *Euglena gracilis* chloroplast low molecular weight RNA was fractionated on 6% and 15% polyacrylamide gels and transferred to nylon membranes (13). These blots were hybridized to RNA probes complementary to unspliced (pEZC2003) and partially spliced *rps18* intron 2 (pEZC1056). The results of these hybridization experiments are shown in Fig. 6. As expected, when fractionated on a 6% polyacrylamide gel, both probes hybridize to RNA species of approximately 105–110 nt which migrate faster than the 123 nt 5S rRNA control. However, when the RNA is fractionated on a 15% polyacrylamide gel, the probes hybridize to RNA species which migrate slower than the 5S rRNA. This shift in mobility is consistent with the hypothesis that excised group III introns accumulate as lariats rather than linear RNA molecules (13). No abundant RNA species larger than 110 nt were detected with the *rps18* intron 2 probes, although faint signals of 317 nt and 261 nt are detected by both *rps18* intron 2 probes on the 6% gel (Fig. 6, lanes 2 and 3). These higher molecular weight RNAs might be splicing intermediates containing intron lariats and 3'-exons. Alternatively these RNAs

Figure 4. Sequence ladders of cloned cDNAs from which both internal introns 2c and 2d have been excised. Plasmid names are listed below each sequence ladder. Positions of excision of intron 2d, coordinates 39759–39870, are indicated by hollow arrowheads. Positions of excision of intron 2c are indicated by solid arrowheads. The coordinates of excised nucleotides of intron 2c for each plasmid are pEYC1055, 39604–39709; pEYC1054, 39604–39710; pEYC1056, 39604–39711; pEYC1057, 39604–39712; and pEYC1058, 39602–39712.

might be excised twintrons, or the result of non-specific hybridization to abundant transcripts, since long exposures were required to detect the 110 nt excised intron.

DISCUSSION

The results of phylogenetic analysis are consistent with the concept that most plastid introns are of recent origin and have arisen by the insertion of mobile introns into existing genes (5, 6). The existence of twintrons is further evidence of the mobility of group II and group III introns (11, 12, 13). The model of twintron formation is that as introns insert into the *Euglena* chloroplast genome, they become targets for subsequent intron insertion. A prediction of this model is that as twintrons accumulate, they should also become targets for intron insertion. The characterization of a complex twintron excised as four group III introns is evidence that introns insert into twintrons and supports the model of mobile introns inserting into the *Euglena* chloroplast genome.

An interesting aspect of group III twintrons and the *rps18* complex twintron is the utilization of multiple splice sites in the excision of some of the group III introns. Three of the four *Euglena* chloroplast group III twintrons use multiple splice sites (13), *rps18* intron 2b uses two 5'-splice sites and *rps18* intron 2c uses two 5'-splice sites and four 3'-splice sites. Why do some group III introns utilize multiple splice sites while others do not?

Figure 5. Positions of internal introns within the proposed structural domains of each of the interrupted introns. Introns 2a and 2b are labelled at their 5'-splice sites. The sites of the internal introns are indicated by arrows. (A) Position of intron 2b within the domain VI-like region of intron 2a. (B) Positions of internal introns 2c and 2d within the domain ID-like and domain VI-like regions of intron 2b, respectively. The nucleotides of potential pairing interaction between the loop of the domain ID-like region of intron 2b and intron 2a are underlined and labelled EBS and IBS, respectively.

There are several possible explanations for the observed phenomena. Since accurate splicing of the external intron is necessary to restore the correct *rps18* reading frame, it is not surprising that *rps18* intron 2a is spliced from invariant boundaries. Perhaps the correct mRNA reading frame is recognized by a mechanism which suppresses mis-splicing of exons. Since the mRNA reading frame is established during translation, the chloroplast 70S ribosome may play a role in the splicing of group III introns (19). The structure of group III introns or the difference in nucleotide composition between introns and exons may be recognized as signals that distinguish exons from introns. Conversely, the use of multiple splice sites could be a property common to group III introns. The frequency of multiple splice site utilization of group III introns interrupting protein coding regions may be similar to that of internal introns of twintrons. It is possible that these events have not been detected because mis-spliced mRNAs are rapidly degraded. Another possibility is that utilization of multiple splice sites is due to mutations in individual introns which alter splice site selection. A mutation in an intron interrupting a protein coding region which results in an altered polypeptide product would be subject to selection and either fixed or lost. But mutations in internal introns of twintrons may be subject to different selection pressures. If a mutation in an internal intron does not prevent excision of the external intron it would be essentially neutral and could persist.

Table 1. Features of *rps18* complex twintron group III introns

<i>rps18</i> Intron	Coordinates	5' Sequence	Size (nt) ^a	3' Sequence	Frequency
2a	39466–39899	UUUUG	107	<u>GAAAAA</u> CAUAUAUA*UUCAAAA	8/8
2b	39562–39888	UUAUG	109	<u>UUACUA</u> CAUUUUA*UAAAAG	5/7
2b	39537–39888	UUUUG	134	<u>UUACUA</u> CAUUUUA*UAAAAG	2/7
2c	39604–39709	UAGUG	106	<u>AAAAAC</u> AUUUAGUUA*UUUUUA	4/15
2c	39604–39710	UAGUG	107	<u>AAAAAC</u> AUUUAGUUA*UUUUUAA	3/15
2c	39604–39711	UAGUG	108	<u>AAAAAC</u> AUUUAGUUA*UUUUUAAU	3/15
2c	39604–39712	UAGUG	109	<u>AAAAAC</u> AUUUAGUUA*UUUUUAAUA	3/15
2c	39602–39711	AAUAG	110	<u>AAAAAC</u> AUUUAGUUA*UUUUUAAU	1/15
2c	39602–39712	AAUAG	111	<u>AAAAAC</u> AUUUAGUUA*UUUUUAAUA	1/15
2d	39759–39870	GUAAG	112	<u>AUUUU</u> AAGAUAUA*UAUUUAUG	8/8
Group III Consensus	N/A	NUNNG	93–119	a b c d e f (3–6N) f'e'd'A* c'b'a' 4N	N/A

^a Sizes of introns (2a) and (2b) assume excision of internal introns from most frequently used splice sites.

In the *rps18* complex twintron there is evidence of two different types of alternate splice site utilization. Intron 2b uses two different 5'-splice sites which are 25 nt apart. In this case, it appears two different splice sites are recognized. When the splice site at position 39537 is used, introns 2a and 2b are 82 nt and 134 nt, respectively. These sizes are outside the normal group III intron range of 90–120 nt. The unusual sizes of the resulting introns may explain the less frequent use of this 5'-splice site (see Table 1). It is not known if the short version of intron 2a resulting from excision of intron 2b from the site at position 39537 can be processed further. Indeed without an *in vitro* system there is no way to determine if any of the observed intermediates can be completely processed. The observance of the splicing of intron 2b at position 39537 as an intron of 134 nt is consistent with flexibility in group III splicing and subsequent splicing of the 82 nt intron 2a is quite possible.

Another pattern of splice site selection is seen in the splicing of intron 2c. In this case, it appears that the correct 3'-splice site is recognized, but multiple adjacent sites are used (see Table 1). Perhaps the observed 3'-splice sites are used with equal frequency because they all have the same domain VI-like structure. However, it appears the mechanism which sets the distance of the 3'-splice site from this structure has been disrupted. This may be analogous to the situation in yeast nuclear pre-mRNA introns in which point mutations uncouple 5'-splice site recognition and utilization (20). It is unclear whether the two 5'-splice sites of intron 2c result from incorrect recognition or utilization. There is a strong bias for the UAGUG at position 39604, although neither of the intron 2c 5'-splice sites possess the U at position +2 (see Table 1). Since the putative normal 5'-splice site is less than ideal, the splicing machinery may recognize or utilize the alternative site.

RNA species corresponding to the size of the excised individual introns, but not excised intact twintrons are detected by RNA hybridization (11, 12, 13, and this work). These results are consistent with sequential excision of individual intron from the twintron. Excision of intact twintrons cannot be ruled out on the basis of these hybridization experiments alone, because the excised intact twintrons may be unstable. However, the detection of partially spliced cDNAs by PCR analysis is evidence that a major *in vivo* RNA maturation pathway of *rps18* intron 2 involves sequential excision of the individual introns. For the individual introns of a twintron to be sequentially spliced, it may be necessary that a functional domain of the external intron be

Figure 6. Northern blot analysis of the excised introns of the *rps18* complex twintron. Isopropanol soluble *Euglena* chloroplast RNA was fractionated on 6% and 15% polyacrylamide slab gels containing 8M urea and transferred to nylon membranes. These blots were hybridized to uniformly ³²P labelled RNA probes complementary to *Euglena* 5S RNA, lanes 1 and 4, the unspliced *rps18* complex twintron, pEZC2003, lanes 2 and 5, and a partially spliced *rps18* cDNA containing only introns 2a and 2b, pEZC1056, lanes 3 and 6. The positions of the *rps18* intron 2 probes are indicated by thick lines below the maps of the recombinant cDNA plasmids.

interrupted (10, 11, 12). For this reason introns-within-introns resemble naturally occurring insertion mutations within cis-acting domains necessary for splicing. Insertions of internal introns within the domain VI-like structures of external group III introns is interpreted as evidence that a domain VI is necessary for group III intron excision. One interpretation of the differences in splice site selection between introns 2c and 2d is that precise excision of intron 2d near the putative unpaired A residue of the domain VI-like structure is necessary to restore the function of this region. The potential secondary structure resembling domain ID in 2b, which is interrupted by intron 2c, may have less stringent structural requirements and therefore be able to accommodate imprecise excision of intron 2c. The presence of internal intron

2c in this region of *rps18* intron 2b is evidence that these structures may function in excision of some group III introns.

We have previously proposed that group III introns have evolved from group II introns by loss and/or reduction of functional domains which have been replaced by trans-acting factors (7, 12, 13, 14). The finding that excised group III introns may accumulate as lariats strengthens this position. Sharp (21) has proposed that nuclear pre-mRNA introns could have descended from group II introns by the evolution of the group II functional domains into the snRNAs of the spliceosome. This fragmentation of introns into trans-acting molecules may allow additional levels of gene regulation and impart genetic diversity (22). Whereas group II intron excision is constrained to specific 5'- and 3'-splice sites by the structure of the intron, trans-acting factors can activate or inactivate alternative splice sites of nuclear pre-mRNAs (23). Utilization of multiple splice sites in the excision of internal introns 2b and 2c of the *rps18* complex twintron would not result in alternative polypeptides, but is evidence that alternative splice site recognition and utilization occurs during splicing in *Euglena* chloroplasts. During the evolution of nuclear pre-mRNA introns from group II introns splicing fidelity has been traded for added flexibility (22). If group III introns have descended from group II introns, they have added variability but the flexibility of multiple polypeptides from a single gene has not yet been detected. It is possible that twintron excision utilizing multiple 5'- and 3'-splice sites is a precursor to alternative splicing which results in multiple polypeptides (11, 12, 13).

The results of our studies of RNA processing in *Euglena gracilis* chloroplasts have several implications concerning intron biology. Both group II and group III introns are likely to be mobile genetic elements which insert into not only coding regions but other introns, twintrons, intergenic spacers (19), and perhaps any RNA. Insertion of introns into existing introns is a mechanism for the assembly of introns with multiple 5'- and 3'-splice sites. Finally, the potential for regulated, alternative splicing may have co-evolved with the fragmentation of group II introns into trans-acting molecules in both group III and nuclear pre-mRNA introns.

ACKNOWLEDGMENTS

We thank Donald Copertino for the *Euglena gracilis* sRNA blots. We also thank Carol Dieckman, Roy Parker, Don Bourque, Donald Copertino and Jennifer Stevenson for helpful comments. This work was supported by grants from the U.S. National Institutes of Health (GM 35625 and 35665 to R.B.H.).

REFERENCES

1. Hallick, R.B. and Stutz, E. (unpublished). X70810.
2. Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) *Mol. Gen. Genet.*, **217**, 185–94.
3. 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.
4. Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., and Ozeki, H. (1988) *J. Mol. Biol.*, **203**, 281–98.
5. Gingrich, J.C. and Hallick, R.B. (1985) *J. Biol. Chem.*, **260**, 16156–16161.
6. Palmer, J.D. (1991) In Bogorad, L. and Vasil, I.K. (eds.), *Molecular Biology of Plastids*. Academic Press, San Diego, pp. 5–53.
7. Christopher, D.A. and Hallick, R.B. (1989) *Nucleic Acids Res.*, **17**, 7591–7608.
8. Siemeister, G., Buchholz, C. and Hachtel, W. (1990) *Mol. Gen. Genet.*, **220**, 425–432.
9. Michel, F., Jacquier, A. and Dujon, B. (1982) *Biochimie*, **64**, 867–881.
10. Michel, F., Umesono, K. and Ozeki, H. (1989) *Gene*, **82**, 5–30.
11. Copertino, D.W. and Hallick, R.B. (1991) *EMBO J.*, **10**, 433–442.
12. Copertino, D.W., Christopher, D.A. and Hallick, R.B. (1991) *Nucleic Acids Res.*, **19**, 6491–6497.
13. Copertino, D.W., Shigeoka, S. and Hallick, R.B. (1992) *EMBO J.*, **11**, 5041–5050.
14. Drager, R.G. and Hallick, R.B. (1993) *Curr. Genet.*, **23**, 271–280.
15. Hallick, R.B., Richards, O.C. and Gray, P.W. (1982) In Edelman, M., Hallick, R. B. and Chua, N.-H. (eds.), *Methods in chloroplast molecular biology*. Elsevier Biomedical, New York, pp. 281–294.
16. Yepiz-Plascencia, G.M., Radebaugh, C.A. and Hallick, R.B. (1990) *Nucleic Acids Res.*, **18**, 1869–1878.
17. Holton, T.A. and Graham, M.W. (1991) *Nucleic Acids Res.*, **19**, 1156.
18. Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
19. Stevenson, J.K., Drager, R.G., Copertino, D.W., Christopher, D.A., Jenkins, K.P., Yepiz-Plascencia, G. and Hallick, R.B. (1991) *Mol. Gen. Genet.*, **228**, 183–192.
20. Parker, R. and Guthrie, C. (1985) *Cell*, **41**, 107–118.
21. Sharp, P.A. (1991) *Science*, **254**, 663.
22. Guthrie, C. (1991) *Science*, **253**, 157–163.
23. Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989) *Annu. Rev. Genet.*, **23**, 527–577.