RNA (1995), *1*:624–633. Cambridge University Press. Printed in the USA.

Alternative splicing of the *Euglena gracilis* chloroplast *roaA* transcript

KRISTIN P. JENKINS,[1] LING HONG,[3] and RICHARD B. HALLICK[2]

[1] Molecular and Cellular Biology Department, and [2] Biochemistry Department, University of Arizona, Tucson, Arizona 85721, USA
[3] Molecular and Cell Biology Department, University of California, Berkeley, California 94720, USA

ABSTRACT

A novel gene, *roaA* (ribosomal operon-associated gene), encoding a potential RNA-binding protein has been identified in the *rpl23* ribosomal protein operon of the *Euglena gracilis* chloroplast genome. The *roaA* gene is interrupted by one group III and three group II introns. Introns 1 and 2 of *roaA* can be interpreted as a twintron, formed from the insertion of a group II intron into the 5′ splice site of a group III intron. Alternative splicing of the group III intron results in two distinct transcripts encoding proteins of 516 and 514 amino acids. Group III introns may play a role in the generation of alternatively spliced transcripts.

Keywords: alternative splicing; RNA-binding protein; *roaA*; twintron

INTRODUCTION

Euglena chloroplast differs from the chloroplasts of higher plants at the level of transcription, RNA processing (Stevenson & Hallick, 1994), and RNA splicing (Copertino & Hallick, 1993). Due to high intron content, splicing plays a major role in *Euglena gracilis* chloroplast RNA metabolism. There are at least 86 group II introns in the *Euglena* chloroplast genome (Copertino & Hallick, 1993). *Euglena* also contains at least 64 group III introns, which are unique to euglenoid protists (Copertino & Hallick, 1993). Group III introns range in size from 97 to 119 nt, and are believed to be abbreviated versions of group II introns. Group II introns can be represented in a complex secondary structure model of six helical domains (I–VI) radiating from a central core (Michel et al., 1989). Domains I, V, and VI have been shown to play a critical role in the self-splicing of group II introns. Group III introns lack domains II–V, but contain domain VI and may contain domain ID. Domains ID and VI are important for splice site selection and lariat formation during group II intron splicing.

Because group III introns lack domains II–V, the splicing activities performed by these domains may be supplied in *trans*. Several proteins have been reported to assist splicing of group I and II introns (Lambowitz & Perlman, 1990). Because group III introns have been found only in euglenoids, proteins involved in splicing group III introns are also likely to be unique to euglenoid plastids. Genes for three maturase-like proteins have been identified in introns of the *Euglena* chloroplast *psbC* operon (Mohr et al., 1993; Copertino et al., 1994).

Euglena chloroplasts also contain at least 15 twintrons, or introns within introns (Copertino & Hallick, 1993). Examples of twintrons include group II within group II introns, group III within group III introns, and "mixed" twintrons with group II introns internal to group III introns, or vice versa. Complex twintrons with multiple internal introns have also been observed (Drager & Hallick, 1993a; Hong & Hallick, 1994; L. Zhang, K.P. Jenkins, E. Stutz, R.B. Hallick, in prep.). Twintrons are believed to be the result of a mobile intron inserting into another intron. Alternative splicing has been observed at both the 5′ and 3′ internal and external splice sites of some group III twintrons (Copertino et al., 1992; Drager & Hallick, 1993a).

There are only a few examples of alternative splicing of non-nuclear introns. Use of alternative 3′-splice sites in *Podospora anserina nad 1-i4* and *cox1-i7* group I mitochondrial introns is implicated in expression of a discontinuously encoded, intronic open reading frame (orf) (Sellem & Belcour, 1994). Alternative splicing of nuclear pre-mRNAs is an important and widespread feature of regulating gene expression in many metazoans. It has been shown to be involved in regulation of important processes such as sex determination, programmed cell death, and temporally and spatially controlled gene expression (Smith et al., 1989). Alternative splicing may also play a role in evolution by allowing

Reprint requests to: Richard B. Hallick, Biochemistry Department, University of Arizona, Tucson, Arizona 85721, USA; e-mail: hallick@arizona.edu.

genetic diversity without requiring permanent changes at the DNA level (Smith et al., 1989). The potential of alternative splicing of group III introns to mirror alternative splicing of nuclear introns and play an important role in chloroplast genome evolution led us to an in-depth analysis of selected group III splicing events.

We have explored whether splicing of group III introns or twintrons might lead to multiple mature mRNAs from a single pre-mRNA. An example of alternative splicing of a group III intron was found in a novel gene, *roaA*, encoded in the *E. gracilis* chloroplast *rpl23* ribosomal protein operon. *roaA* was initially identified as two potential orfs in the intercistronic region between *rps3* and *rpl16* (Christopher & Hallick, 1990). The presence of these orfs was considered unusual, because the gene content and order of the *rpl23* operon is highly conserved (Lindahl & Zengel, 1986; Tanaka et al., 1986; Michalowski et al., 1990), and in *Escherichia coli* and other chloroplasts, *rps3* and *rpl16* are adjacent (Fig. 1). We have determined the orfs constitute a single gene, *roaA*, which is interrupted by one group III and three group II introns. Alternative splicing of the group III intron 2 of *roaA* pre-mRNA results in two distinct mRNAs encoding polypeptides of either 514 or 516 amino acids. This alternative splicing may have evolved after formation of a twintron by insertion of intron 1 into the 5′ splice site of intron 2. Intron insertion may play an important role in evolution of plastid genes of euglenoid protists via the potential for alternative splicing.

RESULTS

cDNA analysis

The *rps3-rpl16* intercistronic region is co-transcribed with the ribosomal protein genes of the *rpl23* operon (Christopher & Hallick, 1990). Intercistronic processing results in an approximately 1.4 kb RNA containing an *rps3-rpl16* intercistronic transcript. To determine if the orfs were independent or comprise a single gene, cDNAs from the *rps3-rpl16* intercistronic region were studied.

The 1.4 kb RNA species detected in northern analysis might be the spliced mRNA product of a larger pre-mRNA from the 2.8 kb region. To investigate this possibility, RNA from the 2.8 kb region was PCR amplified, cloned, and sequenced. The nucleotide sequence of the 2.8 kb region and the location of primers used for PCR amplification are shown in Figure 2. A cDNA primer (cDNA3) located in the first exon of the *rpl16* gene was used to prime cDNA synthesis of *E. gracilis* chloroplast RNA. Resulting cDNAs were amplified using the PCR primer PCR1 (Fig. 2). The PCR products were cloned and sequenced. A single continuous orf comprised of five exons was detected.

The translation start of the orf was not defined. AUG codons are absent in the genomic sequence between the stop codon of *rps3* and the beginning of the orf. In fact, the first in-frame AUG codon is in exon 5. The orf may use the alternative start codon UUG. UUG, GUG, and AUU are used as alternative start codons in *E. coli* (Gren, 1984) and plastids (Michalowski et al., 1990). For example, in *E. gracilis* chloroplast, the *atpF* gene apparently uses the alternative start codon GUG (Drager & Hallick, 1993b). If the UUG codon 55-nt downstream of *rps3* (coordinates 53736–53738) is used as the initiator codon, a protein of 516 amino acids is predicted. This 516-amino acid orf has been designated *roaA* for ribosomal operon-associated gene, and "A" as the first such gene. The gene structure is exon 1, 49 nt; intron 1, 349 nt; exon 2, 4 nt; intron 2, 97 nt; exon 3, 613 nt; intron 3, 325 nt; exon 4, 18 nt; intron 4, 438 nt; exon 5, 867 nt. The entire gene is 2,760 nt long (Fig. 2) (coor-

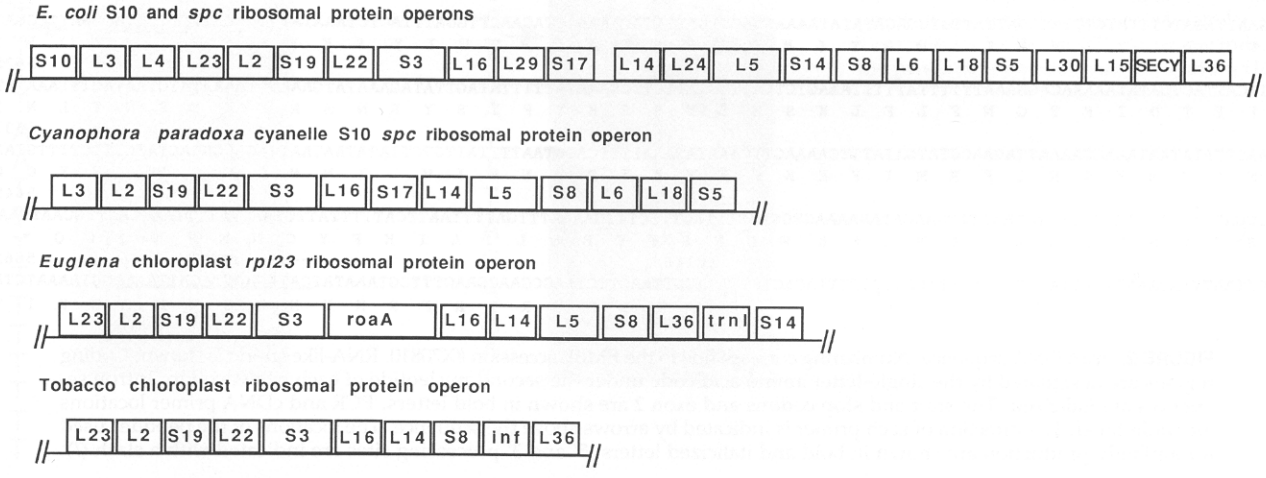

FIGURE 1. Comparison of ribosomal protein operons. The gene content and order of the *E. coli* S10 and *spc* operons are compared with that of the *Cyanophora*, *Euglena*, and tobacco chloroplast ribosomal protein operons.

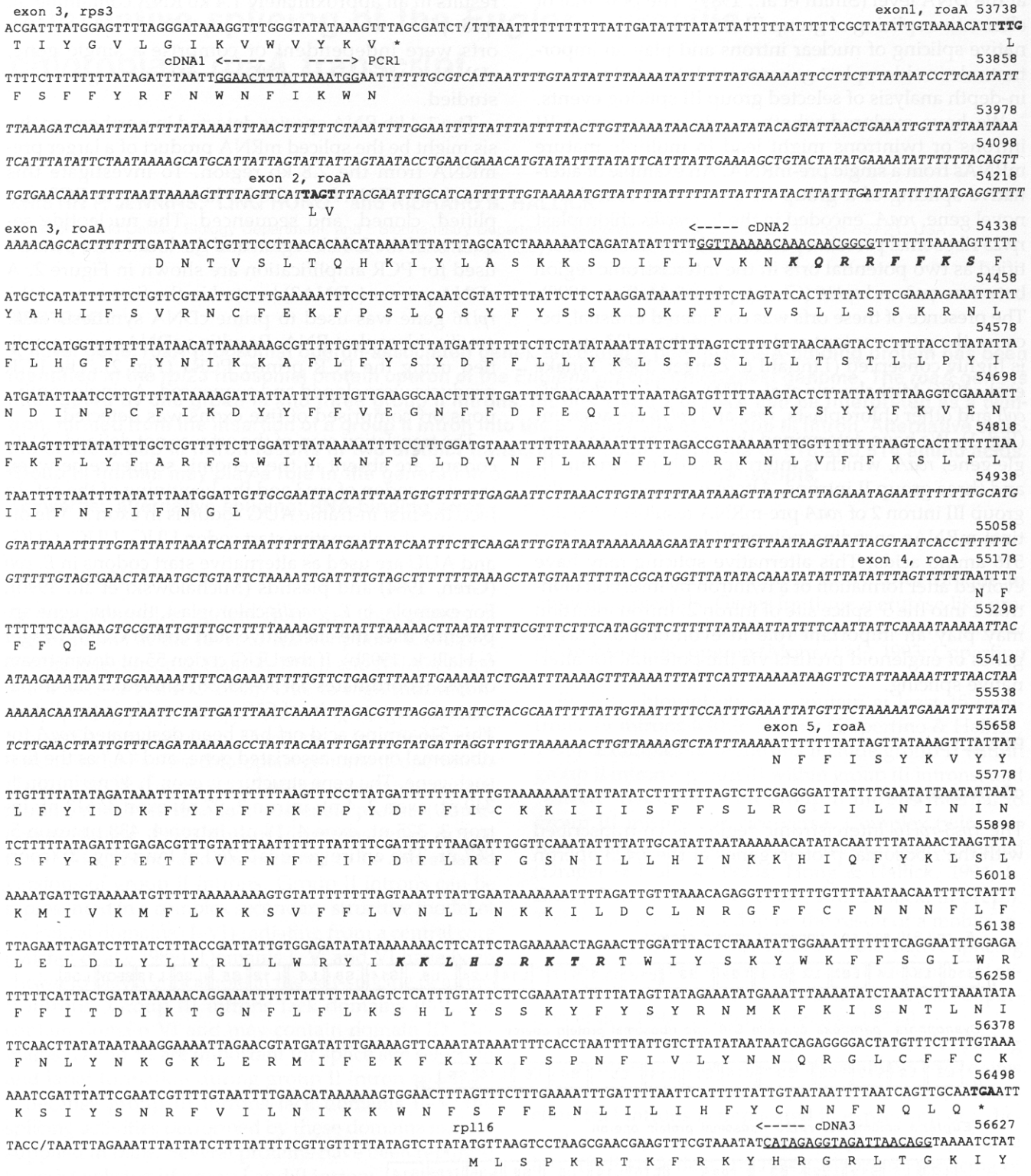

FIGURE 2. *roaA* DNA sequence. Numbering corresponds to the EMBL accession #X70810. RNA-like strand is shown. Coding regions are designated by the single-letter amino acid code under the second nucleotide of each triplet codon. Intron sequences are italicized. The start and stop codons and exon 2 are shown in bold letters. PCR and cDNA primer locations are underlined. The direction of each primer is indicated by arrows above the sequence. The locations of the peptides used for antibody production are shown in bold and italicized letters. 5′ and 3′ processing sites are indicated with a slash (/).

dinates 53736–56495). Introns 1, 3, and 4 are group II introns. Intron 2 is a group III intron. Exon 2, at 4 nt long, is the smallest plastid exon reported to date.

mRNA analysis

Rather than using UUG, it is possible that an AUG start codon is supplied by *trans*-splicing of the roaA transcript. To test this hypothesis, the 5′ end of the *roaA* transcript was analyzed by RNA sequencing (Fig. 3). The primer cDNA1 in exon 1 of *roaA* was ^{32}P-end-labeled and used to prime cDNA synthesis from *E. gracilis* chloroplast RNA by reverse transcriptase. Sequence ladders were generated from cDNA reactions using deoxy- and dideoxynucleotides. The resulting sequence extended 65 (±2) nt upstream of the *roaA* UUG start codon to a major stop in all lanes at position 53673 (±2 nt) (Fig. 3). Minor stops observed 3′ of the major stop may be due to secondary structure. Sequence extending beyond the major stop was the result of unprocessed transcripts containing *rps3* and *roaA*. Although it is possible that the mRNA is *trans*-spliced 3′ of the primer cDNA1, based on this experiment, *roaA* is not *trans*-spliced and begins with the alternative codon UUG.

The 3′ end of the *roaA* mRNA was defined by locating the processing site between *roaA* and *rpl16*. A primer, cDNA3, complementary to exon 1 of *rpl16*, was ^{32}P-end-labeled and used to prime cDNA synthesis from *E. gracilis* chloroplast RNA. A major stop in the sequence at position 56502 (±2 nt), mapped 50 (±2) nt upstream of the *rpl16* start codon and 7 (±2) downstream of the *roaA* stop codon. It is possible that the 3′ end of the *roaA* transcript is processed further follow-

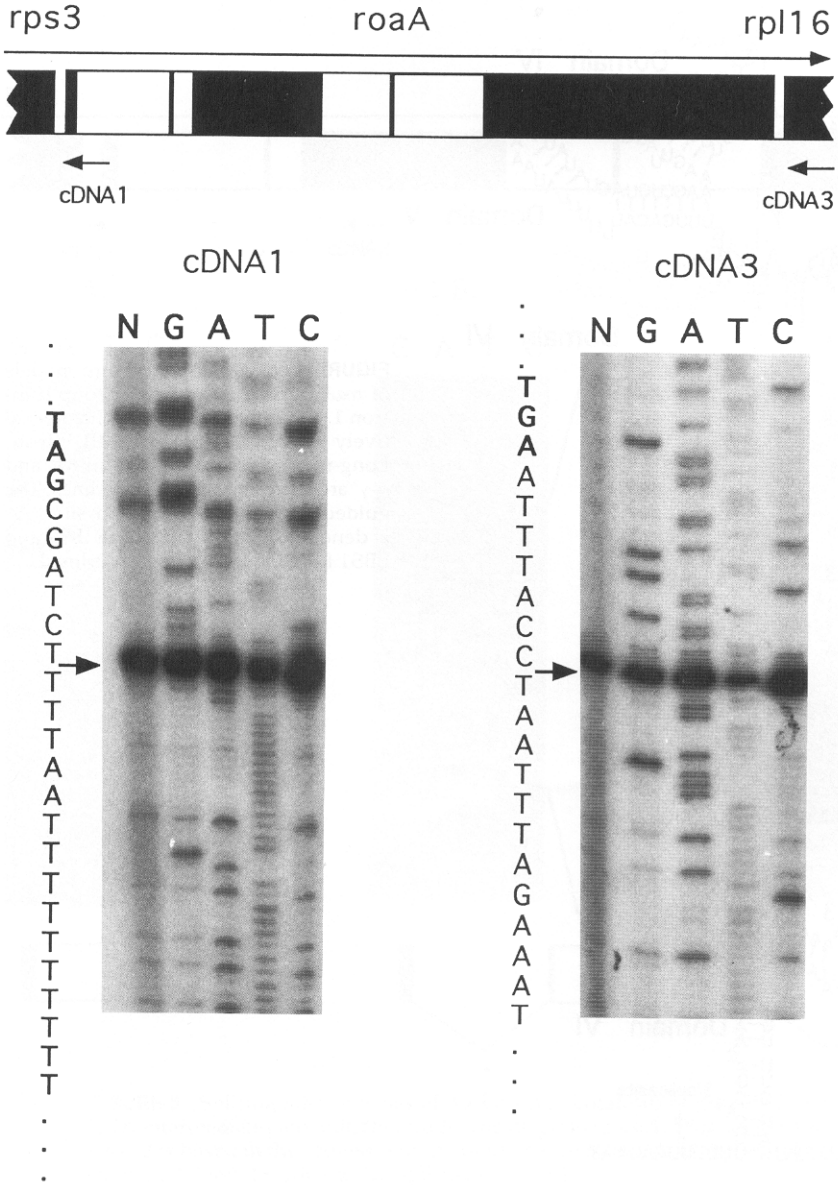

FIGURE 3. Analysis of the 5′ and 3′ ends of the *roaA* transcript. A diagram of the *roaA* gene is shown, and approximate primer locations are indicated by labeled arrows under the diagram. RNA sequence from exon 1 of *roaA* with primer cDNA1 is shown in the left panel. The primer is located in exon 1 of *roaA* (nt 53764–53780). The RNA sequence is indicated to the left of the panel. An arrow indicates the processing site between *rps3* and *roaA*. RNA sequence from exon 1 of *rpl16* with primer cDNA3 is shown in the panel on the right. The primer is located in exon 1 of rpl16 (nt 56588–56607). An arrow indicates the processing site between *roaA* and *rpl16*.

ing intercistronic cleavage. Based on these experiments, the monocistronic, fully spliced, and processed *roaA* transcript is 1,623 (±4) nt.

roaA Introns

The *E. gracilis* chloroplast genome contains a total of 157 group II and group III introns (Copertino & Hallick, 1993). *roaA* is interrupted by one group III and three group II introns. The predicted secondary structures of the *roaA* introns have been compared to other *E. gracilis* group II and group III introns. Introns 1, 3, and 4 can be folded into the typical group II intron secondary structure model, with six stem loops (I–VI) radiating from a central core (Michel et al., 1989). The proposed secondary structure of intron 1 is shown in Figure 4A. The proposed structure is similar to other small group II introns of *Euglena*. In addition to the six

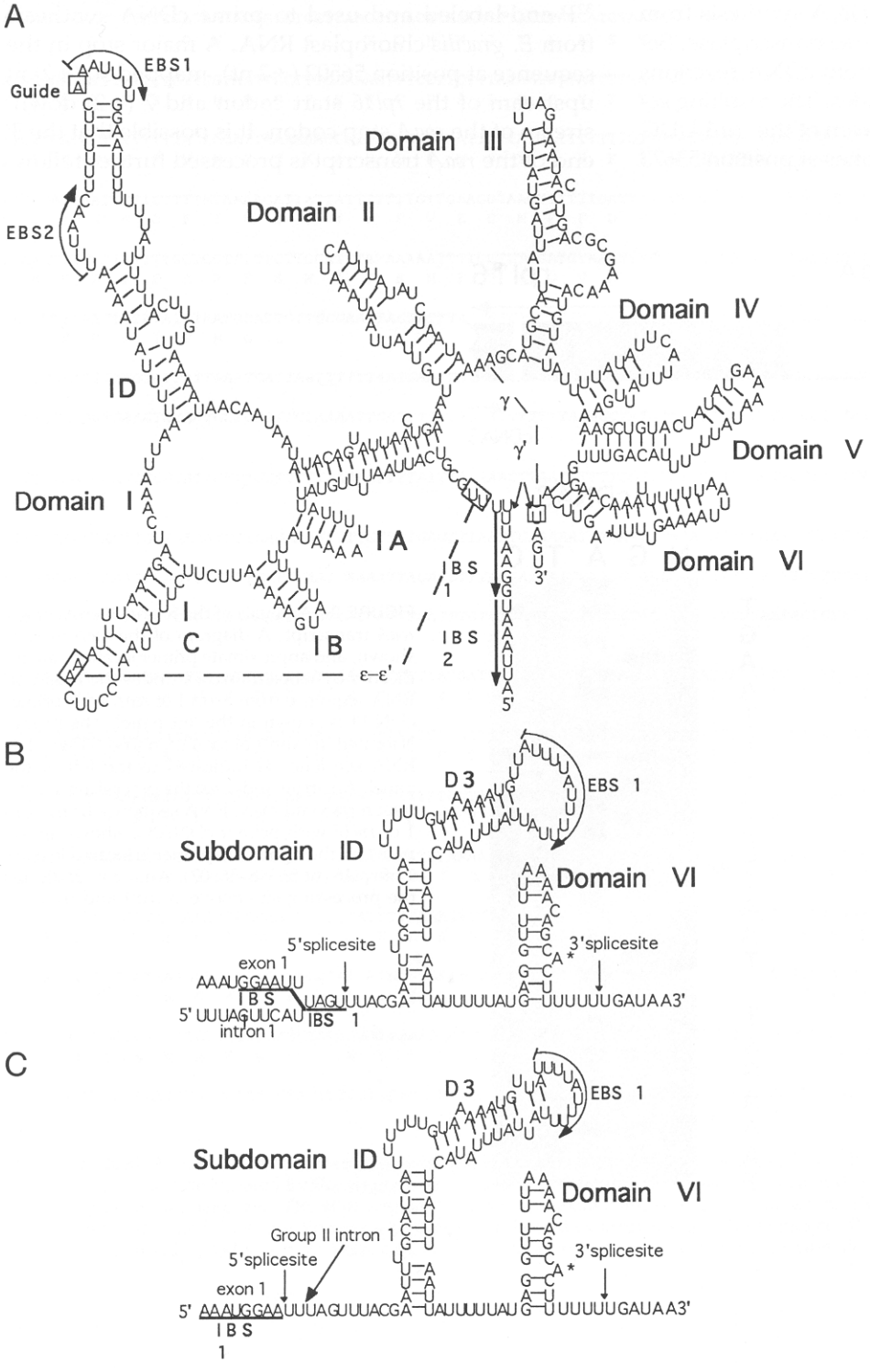

FIGURE 4. Secondary structure models of *roaA* introns 1 and 2. **A:** Group II intron 1. **B:** Group III intron 2. **C:** Alternatively spliced 103-nt group III intron. Long-range tertiary interactions ϵ-ϵ' and γ-γ' are represented by dashed lines. The guided pair is boxed. Branch site "A" is denoted by an asterisk. The IBS1 and EBS1 for each intron are underlined.

helical domains, I–VI, five conserved tertiary interactions, ϵ-ϵ', γ-γ', EBS1-IBS1, EBS2-IBS2, and the guided pair are predicted.

Intron 2 is a 97 nt group III intron. The predicted secondary structure and EBS-IBS of intron 2 is shown in Figure 4B. As with most group III introns, a subdomain ID and domain VI can be identified (Copertino & Hallick, 1993). As shown in Figure 4B, intron 2 has an EBS-like domain that may interact with an IBS partially in exon 1. If this interaction is required for splicing of intron 2, intron 1 must be spliced first to restore the intron 2 IBS.

Alternative splicing of exon 2

To investigate the splicing pattern of introns 1 and 2, RNA splicing intermediates were examined using cDNA and PCR techniques. PCR primers PCR1 and cDNA2 were used to amplify *roaA* cDNAs (Fig. 5). The PCR products were cloned and 23 independent cDNAs were sequenced. Three cDNAs contained a 556 nt unspliced mRNA (Fig. 5A), two contained a 207 nt partially spliced mRNA with intron 1 excised (Fig. 5B), and 16 contained a 109 nt completely spliced mRNA (Fig. 5C). None of the cDNAs contained the partially spliced product with intron 2 removed and intron 1 retained (predicted 458 nt). In addition, the 458 nt partially spliced product was not detected when PCR products were fractionated on agarose gels.

Two of the completely spliced cDNAs represented alternatively spliced mRNAs compared to the 21 other cDNAs (Fig. 5D). In these cDNAs, an alternative intron 2 splice site 2 nt 5′ of the predicted splice is present. Along with intron 2, exon 2 and 2 nt of exon 1 are excised, and the remainder of exon 1 is ligated directly to exon 3. The expected and alternatively spliced

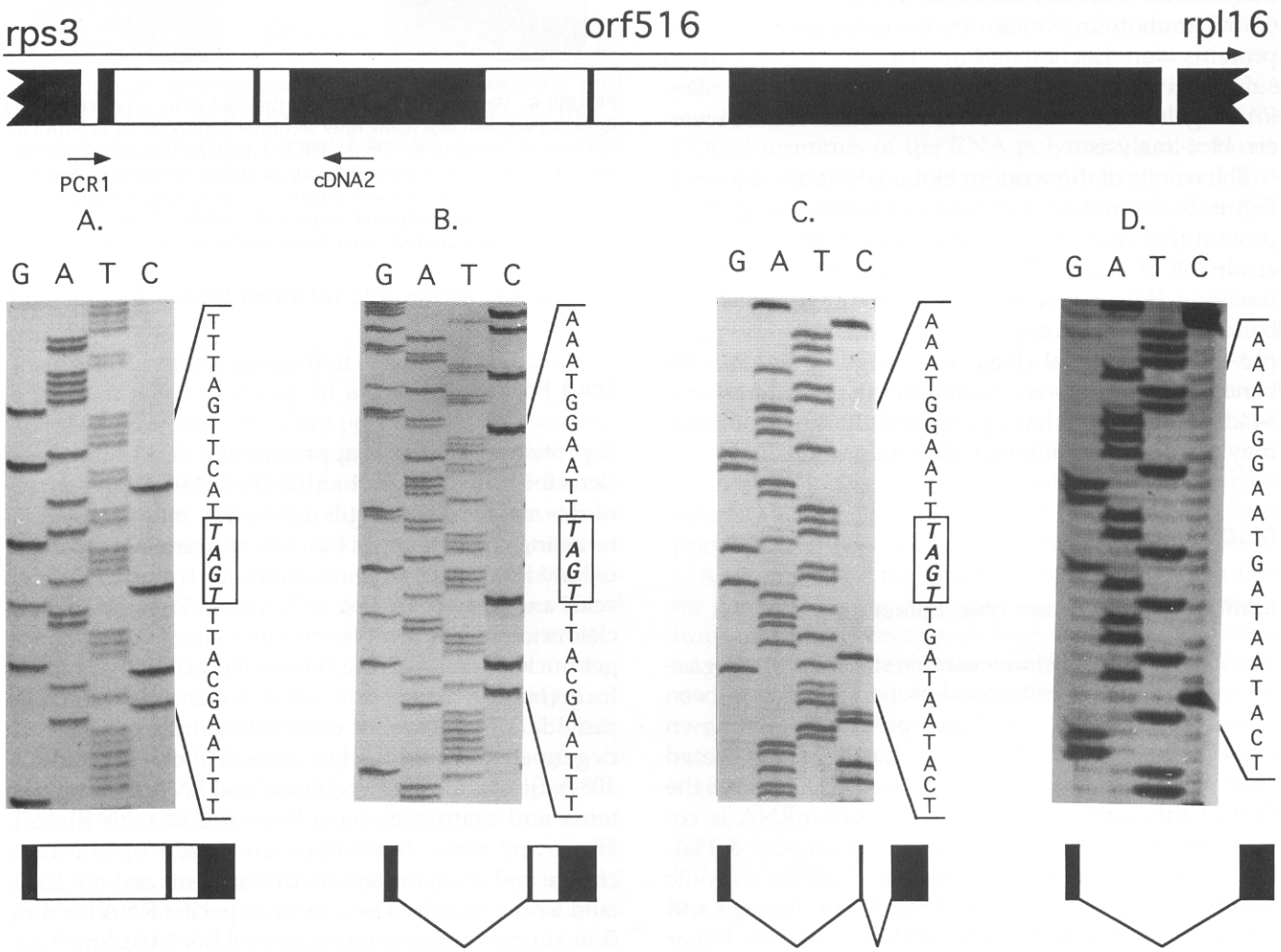

FIGURE 5. Splicing intermediates of *roaA* exon 2. Locations of primers PCR1 and cDNA2 used for PCR amplification of RNA intermediates are indicated under the diagram of *roaA*. Diagrams of the cDNA clones are shown under each panel. Exon 2 is boxed in the sequence to the right of each panel. **A:** Unspliced. **B:** Intron 1 excised. **C:** Introns 1 and 2 excised, exon 2 retained. **D:** Alternative splicing of exon 2.

products differ in length by 6 nt, and can be resolved on 5% nondenaturing acrylamide gels (data not shown). Although the alternatively spliced transcript appears to be less abundant than the predicted transcript, it does accumulate to significant levels. Alternative splicing does not change the reading frame of the protein. However, codons for amino acids 17 and 18 (LEU-VAL) are missing, resulting in an orf of 514 amino acids.

The *roaA* protein product

The *roaA* protein product was identified by western analysis of *E. gracilis* chloroplast proteins. To generate antibodies against *roaA*, two synthetic oligopeptides, KQRRFFKS (amino acids 44–51) and KKLHSRKTR (amino acids 374–382) were made. Chickens were inoculated with either peptide and antibodies were isolated from egg yolks (Polson et al., 1985).

To prepare chloroplast proteins for immunodetection, isolated *E. gracilis* chloroplasts were lysed mechanically by freeze-thawing and Dounce homogenization. Membranes were separated from proteins soluble in 0.5 M ammonium sulfate by centrifugation. Soluble proteins were fractionated in 30 and 70% ammonium sulfate precipitations. Proteins were separated on SDS-PAGE gels, and transferred to nitrocellulose for western blot analysis.

The results of the western blot analysis are shown in Figure 6. Combined anti-*roaA* antibodies recognize a protein that migrates at approximately 80 kDa in the crude 0.5 M ammonium sulfate and 70% precipitant fractions. Both antibodies recognize the same 80 kDa band when used independently (data not shown). A pre-immune control does not recognize the 80 kDa band (data not shown). Although the predicted size is 64 kDa, the protein has a predicted charge of +57 that may result in anomalously slow migration.

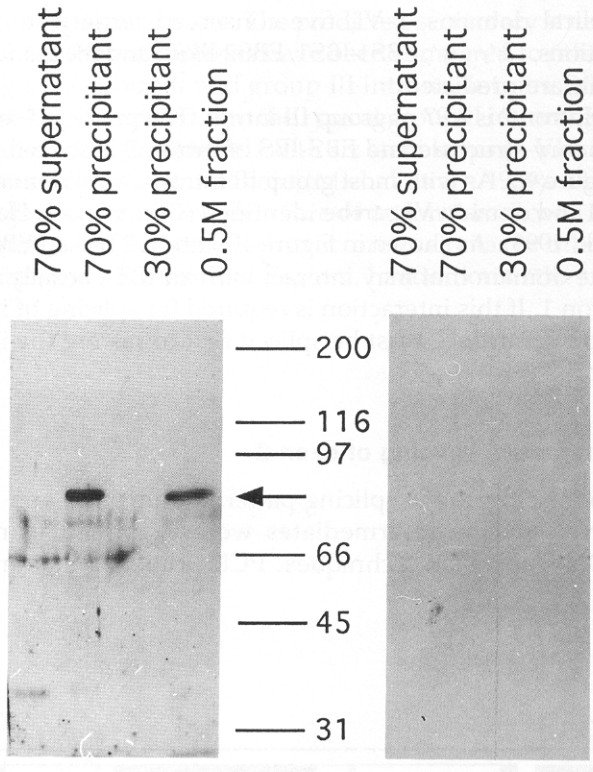

FIGURE 6. Western blot analysis of the *roaA* gene product. On the right is a western blot using only secondary antibody as a control for nonspecific interactions. On the left is a western blot using both anti-*roaA* antibodies and secondary anti-chicken antibody. Immunoreactive bands, indicated with an arrow at approximately 80 kDa, are visible in the 70% precipitant and 0.5 M fractions. Molecular weight marker locations are indicated between the two analyses.

DISCUSSION

Identification of a new chloroplast gene

The 2.3 kb *rps3-rpl16* intercistronic spacer of the *E. gracilis* chloroplast *rpl23* ribosomal protein operon has been found to encode a novel and previously unknown chloroplast gene. The new gene has been designated "roa" for "ribosomal operon-associated" and "A" as the first of this class. The 2.3 kb *roaA* pre-mRNA is co-transcribed with the *rpl23* operon (Christopher & Hallick, 1990), and then processed to a mono-cistronic 1.6 kb mRNA by 5′ and 3′ endonuclease cleavage and splicing of four introns. The mRNA is slightly larger than the previous estimate of 1.4 kb, which was based on northern hybridization. The predicted protein product is a basic polypeptide of 516 amino acids. This polypeptide is expressed in *E. gracilis* chloroplast.

roaA Protein

A protein migrating at approximately 80 kDa has been detected with anti-*roaA* antibodies. Although the role of the *roaA* protein is still unknown, it has several interesting characteristics that are indicative of an ability to bind RNA. The protein contains 28% aromatic amino acids and has a charge of +57. Aromatic residues in nucleic acid binding sites of proteins can intercalate in target nucleic acids, forcing the nucleic acid into a specific form (Helene & Lancelot, 1982). A charge of +57 could provide a high level of electrostatic interactions with negatively charged nucleic acids (Helene & Lancelot, 1982). Similarities between *roaA* and RNA binding proteins and maturases have been found with BLAST. However, these similarities are based on the high charge and aromatic amino acid content, and not localized to any motifs. The lack of a specific RNA binding domain may reflect a more general RNA binding function. The location of *roaA* in a ribosomal protein operon is also compatible with a potential role as an RNA binding protein involved in maturation or translation of mRNA.

Group III intron splice site selection

An unusual feature of the *roaA* gene structure underscores an interesting splicing problem. The 5′ splice sites of group II introns are defined by tertiary interactions between the EBS and IBS (Michel et al., 1989). Because group III introns are believed to be degenerate group II introns, the same mechanism for 5′ splice site selection has been proposed (Copertino et al., 1994). However, exon 2 of *roaA* is only 4 nt long, which may be too small to contain the IBS required for splicing of the downstream group III intron (intron 2). The two other sites that could complete the IBS for intron 2 are within either intron 1, or exon 1 after excision of intron 1. As shown in the model for group III intron structures in Figure 4B, the putative EBS1-IBS1 pairing is more extensive if part of IBS1 is derived from exon 1 (5′GGAAUUUAGU) rather than intron 1 (5′UUCA UUAGU).

Splicing of introns 1 and 2 would normally be expected to be unordered (Hong & Hallick, 1994). However, the splicing intermediate containing intron 1 and lacking intron 2 has not been detected. Although the existence of this intermediate cannot be ruled out, the predominant intermediate species lacks intron 1 and retains intron 2. This example of ordered splicing is consistent with the proposed model in which EBS–IBS interactions function in 5′ splice site selection in group III introns.

Possible structural basis for alternative splicing

Intron 2 is a single intron that can be spliced as either a 97 nt or a 103 nt group III intron, depending on 5′ splice site selection. We have previously suggested that group III introns might have a tertiary interaction comparable to the EBS1-IBS1 pairing of group II introns (Copertino et al., 1994). As suggested in the models in Figure 4C, different EBS1–IBS1 tertiary interactions are possible for the two alternatively spliced variants of intron 2. The 103 nt intron is spliced when exon 2 and 2 nt of exon 1 are excised with intron 2. In Figure 4C, subdomain ID for the 103 nt intron is shown to have a different EBS than the 97 nt intron. The 103 nt intron EBS–IBS interaction shifts the 5′ splice site to UUUAG. This splice site fits the group III consensus sequence NUNNG. However, the UUUAG alternative splice site cannot be used if intron 1 is present, because intron 1 is located between the second and third uridines in this splice site. In the 103 nt intron splicing pathway, intron 1 is inserted into the 5′ splice site of intron 2 and must be spliced to restore the splicing ability of intron 2. Thus, intron 1 and the 103 nt version of intron 2 represent a new twintron, the 16th in the *E. gracilis* genome. This twintron is only the second example of a group II intron within a group III intron. The first example is in *rps3*, the upstream gene (Copertino et al., 1992).

From an evolutionary standpoint, the existence of this twintron means that the "alternatively" spliced species lacking exon 2 is actually the ancestral species. Exon 2 may have evolved from the 5′ boundary of the interrupted ancestral group III intron. To test this hypothesis, we are investigating the intron content of *roaA* in more ancient *Euglena* species. The prediction is that *Euglena* species lacking introns 1 and 2 would encode a single mRNA of 514 amino acids.

Alternative splicing of intron 2

The evolution of exon 2 from a group III intron would support the theory that alternative splicing of group III introns has played a role in the evolution of the *Euglena* chloroplast genome. Alternative splicing of *E. gracilis* chloroplast *rpl16* and *rpoC* group III introns has been described previously (Copertino et al., 1992). In *rpl16* and *rpoC*, the internal group III introns of group III twintrons are excised using alternative 5′ and 3′ splice sites. In *rpl16*, the product is a pre-mRNA that, if translated, would yield a truncated *rpl16* protein. In *rpoC*, translation of the alternatively spliced pre-mRNA would result in the addition of four amino acids to the COOH-terminus of the RNA polymerase β'-subunit. It is not known if these alternatively spliced pre-mRNAs are translated or if they undergo further processing to result in the fully spliced transcript.

The alternative splicing of *roaA* intron 2 is the first example of exon skipping during plastid mRNA processing. Although the two distinct polypeptide products of *roaA* of 514 and 516 amino acids were not resolved by western blot analysis and have not yet been confirmed by amino terminal sequence analysis, it is reasonable to speculate that both mature mRNAs are translated. Exon skipping is a common feature of alternatively spliced nuclear pre-mRNAs, but has been associated previously only with nuclear spliceosomal reactions.

Alternative splicing of introns is another mechanism for generation of genetic diversity. Introns and twintrons may be so prevalent in *Euglena* because they allow adaptability via alternative splicing. Group III introns are good candidates for evolutionary mediators of alternative splicing because splice site selection is less constrained by secondary structure and tertiary interactions than in group I and group II introns.

On the evolutionary origin of *roaA*

Because *roaA* is absent from all other known plastid *rpl23* operons, it is possible that the *roaA* gene was inserted in the *E. gracilis* plastid genome after evolutionary descent from a common ancestor with other photosynthetic eukaryotes. The *rpl23* operon of *E. gracilis* is unique among known plastid *rpl23* operons in be-

ing interrupted by 9 group II and 15 group III introns, including two twintrons. From an evolutionary analysis of intron content and position in various species of *Euglenaphycea*, it was concluded that introns in *E. gracilis* are a derived characteristic, having been added to genes that have evolved from intron-less progenitor genes (M.D. Thompson, D.W. Copertino, E. Thompson, M.R. Favreau, R.B. Hallick, in prep.). The origin of the *roaA* gene may be associated with the invasion of group II and group III introns into the *rpl23* operon.

A noteworthy feature of the *rpl23* operon is a series of intercistronic, group III introns (Stevenson et al., 1991). One hypothesis is that *roaA* was first introduced into the genome as an intron-encoded maturase from a mobile group III intron that inserted into the *rps3-rpl16* intercistronic spacer. The maturase may have evolved into a more general RNA-binding protein as its function was replaced by *trans*-acting proteins. The group III intron features may also have been lost through genetic drift. A precedent for this hypothesis is found in *ycf13*, a group III intron-encoded maturase-like protein, in intron 4 of the photosystem II *psbC* gene of *E. gracilis* (Copertino et al., 1994). Although the *psbC* gene has been lost from the plastid genome of the non-photosynthetic protist *Astasia longa* (Siemeister et al., 1990), the *ycf13* gene has been retained as a free-standing cistron. Presumably, *ycf13* was an intron-encoded orf in the mutual ancestor of both *E. gracilis* and *Astasia longa*. Whatever role *ycf13* plays, it has been conserved in *Astasia* despite the loss of its original carrier intron. In a like fashion, *roaA* may have been maintained in *Euglena* as its carrier intron was lost.

MATERIALS AND METHODS

The complete *E. gracilis* chloroplast genome is available in Genebank, EMBL accession #X70810. All coordinates in the manuscript refer to release 42, version 36 of this accession number.

RNA isolation

Chloroplasts were isolated from photoautotrophically grown *E. gracilis* as described (Hallick et al., 1982). RNA was isolated from purified chloroplasts as described previously (Stevenson & Hallick, 1994). Briefly, isolated chloroplasts were phenol-chloroform extracted in NTES buffer (100 mM NaCl; 10 mM Tris-HCl, pH 7.5, 1 mM EDTA; 1% SDS) and the nucleic acid was ethanol precipitated. DNA was removed by treating with RQ1 DNase (1 U/1 μg, Promega) in the presence of RNasin (40 U/10 U DNase, Promega) followed by a final phenol-chloroform extraction.

cDNA synthesis, amplification, and cloning

cDNAs were synthesized using specific oligonucleotide primers (University of Arizona Biotechnology Center). For cDNA synthesis reactions, 200 ng of cDNA primer was used to prime cDNA synthesis from 5 μg of chloroplast RNA using the BRL cDNA synthesis kit. Primer cDNA1 (5′-CCATTTAA TAAAGTTCC-3′) and its complement PCR1 (5′-GGAACTTT ATTAAATGG-3′, coordinates 53764–53780), are in exon 1. Primer cDNA2 (5′-CGCCGTTGTTTGTTTTTAACC-3′, coordinates 54301–54321) is in exon 3. Primer cDNA3 (5′-CCTG TTAATCTACCTCTATG-3′, coordinates 56588–56607) is in exon 1 of *rpl16*. Amplification programs consisted of 25 cycles with a dissociation segment at 94 °C for 1 min, an annealing segment at 5 °C below the melting temperature of the primer (42–56 °C) for 1.5–2 min, and an extension segment at 72 °C for 2–3 min. PCR products were cloned in pBS+ (Vector Cloning Systems), pKS- (Stratagene), or the TA cloning vector (Invitrogen). Cloned cDNAs were sequenced by standard methods using the Sequenase DNA sequencing kit (U.S. Biochemical).

RNA primer extension

cDNA primers cDNA1 and cDNA3 were 5′ end-labeled using 20 units of T4 polynucleotide kinase (Promega). For each ^{32}P-labeled primer, 1×10^7 cpm was precipitated with 10 μg of RNA. The RNA–primer mix was resuspended in 12 μL annealing buffer (200 mM KCl, 10 mM Tris-HCl, pH 8.3 at 42 °C) and annealed for 2.5 h at 42 °C. The primer extension reactions were carried out in the presence or absence of ddNTPs as described (Christopher & Hallick, 1989).

Computer analysis

Various search and comparison algorithms from the University of Wisconsin GCG package (Devereux et al., 1984) were used in analysis of both DNA and protein sequences. Amino acid and nucleic acid analyses were performed at NCBI using the BLAST network service (Altschul & Lipman, 1990) and the protein analysis program BLOCKS (Henikoff & Henikoff, 1991). Additional searches of the Swissprot, Prosite, and PDB databases were performed at the GenQuest server using the Smith–Waterman comparison program (Smith & Waterman, 1981). The DNA Strider program (Christian Marck) was used for additional sequence analysis with a Macintosh II.

Protein preparation

Euglena chloroplasts were isolated as described previously (Hallick et al., 1982). Isolated chloroplasts were Dounce homogenized in TE buffer (10 mM Tris, 1 mM EDTA) and 1 mM AEBSF, a protease inhibitor (Calbiochem). $(NH_4)_2SO_4$ was added to 0.5 M and the membranes were removed by centrifugation at approximately 225,000 $\times$ *g* for 60 min. The supernatant was collected and 30 and 70% $(NH_4)_2SO_4$ precipitations were performed. The fractions were desalted in Centriprep 30s and concentrated in Centricon 30s (Amicon). Samples were run on 10% SDS-PAGE gels and a wet electrophoretic transfer to nitrocellulose was done for western analysis in standard transfer buffer (20% methanol, 25 mM Tris, 192 mM glycine).

Immunodetection of the roaA gene product

Antigenic peptides were selected using the UWGCG PeptideStructure program (Jameson & Wolf, 1986). These pep-

tides were synthesized as multiple antigenic peptides (MAPs) (Schaaper et al., 1989) at the University of Arizona Biotechnology Center. A lysine core acts as a carrier protein in MAPs and the antigenic peptide is assembled on this core, eliminating the need for coupling to a carrier protein. Fifty micrograms of MAP in Freund's complete or incomplete adjuvant were used to inoculate white leghorn chickens. Antibodies against the individual peptides were raised separately by inoculating chickens with only one antigen. As a pre-immune control, one chicken was inoculated and boosted with Freund's adjuvant alone. Eggs were collected and stored at 4 °C. Antibodies were isolated from the yolks according to Polson et al. (1985). Approximately three yolks were used for each antibody preparation. Yolk proteins were removed from isolated yolks by 3.5% PEG precipitation in PBS (137 mM NaCl, 0.3 mM KCl, 1 mM Na_2HPO_4, 0.2 mM KH_2PO_4, pH 7.2–7.4 (Sambrook et al., 1989)). IgY antibodies were isolated from the supernatant by 12% PEG precipitation. A subsequent 50% ethanol extraction was done to remove the PEG. The antibodies were dialyzed overnight against buffer P (PBS without NaCl) and stored at 4 °C or −70 °C. Primary chicken antibodies were used at a 1:1,000 dilution. A polyclonal rabbit anti-chicken IgG peroxidase conjugated antibody (Sigma) was used as a secondary antibody. To remove nonspecific interactions between the secondary antibody and *Euglena* chloroplast proteins, secondary antibodies were incubated at a 1:100 dilution with *Euglena* chloroplast proteins immobilized on nitrocellulose. The supernatant was collected and used at a 1:100 dilution in western analysis. The ECL western detection kit (Amersham) was used for immunodetection experiments.

ACKNOWLEDGMENTS

We thank Drs. D. Copertino, J. Stevenson, M. Jenkins, K. Moore, and K. Oishi for critical reading of this manuscript. We also thank Drs. C. Hibbert, G. Wildner, and S. Shigeoka for helpful discussions and Dr. C. Frey for help with immunodetection. This work was supported by NIH grant #35665.

Received June 23, 1995; returned for revision July 13, 1995; revised manuscript received July 21, 1995

REFERENCES

Altschul SF, Lipman DJ. 1990. Protein database searches for multiple alignments. *Proc Natl Acad Sci USA 87*:5509–5513.

Christopher DA, Hallick RB. 1989. *Euglena gracilis* chloroplast ribosomal protein operon: A new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Res 17*:7591–7608.

Christopher DA, Hallick RB. 1990. Complex RNA maturation pathway for a chloroplast ribosomal protein operon with an internal tRNA cistron. *Plant Cell 2*:659–671.

Copertino DW, Hall ET, Van Hook FW, Jenkins KP, Hallick RB. 1994. A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res 22*:1029–1036.

Copertino DW, Hallick RB. 1993. Group II and group III introns of twintrons: Potential relationships to nuclear pre-mRNA introns. *Trends Biochem Sci 18*:467–471.

Copertino DW, Shigeoka S, Hallick RB. 1992. Chloroplast group III twintron excision utilizing multiple 5′- and 3′-splice sites. *EMBO J 11*:5041–5050.

Devereux J, Haeberli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res 12*: 387–395.

Drager RG, Hallick RB. 1993a. A complex twintron is excised as four individual introns. *Nucleic Acids Res 21*:2389–2394.

Drager RG, Hallick RB. 1993b. A novel *Euglena gracilis* chloroplast operon encoding four ATP synthase subunits and two ribosomal proteins contains seventeen introns. *Curr Genet 23*:271–280.

Gren EJ. 1984. Recognition of messenger RNA during translation initiation in *Escherichia coli*. *Biochemie 66*:1–29.

Hallick RB, Richards OC, Gray PW. 1982. Isolation of intact, superhelical chloroplast DNA from *Euglena gracilis*. In: Edelmann M, Hallick RB, Chua NH. *Methods in chloroplast molecular biology*. New York: Elsevier Biomedical. pp 281–294.

Helene C, Lancelot G. 1982. Interactions between functional groups in protein–nucleic acid associations. *Prog Biophys Mol Biol 39*:1–68.

Henikoff H, Henikoff JG. 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Res 19*:6565–6572.

Hong L, Hallick RB. 1994. A group III intron is formed from domains of two individual group II introns. *Genes & Dev 8*:1589–1599.

Jameson B, Wolf H. 1986. PEPTIDESTRUCTURE in the GCG package. *CABIOS 4*:181–186.

Lambowitz AM, Perlman PS. 1990. Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci 15*:440–444.

Lindahl L, Zengel JM. 1986. Ribosomal genes in *Escherichia coli*. *Annu Rev Gen 20*:297–326.

Michalowski CB, Pfanzagl B, Löffelhardt W, Bohnert HJ. 1990. The cyanelle S10 spc ribosomal protein gene operon from *Cyanophora paradoxa*. *Mol Gen Genet 224*:222–231.

Michel F, Umesono K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns—A review. *Gene 82*:5–30.

Mohr G, Perlman PS, Lambowitz AM. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res 21*:4991–4997.

Polson A, Coetzer T, Kruger J, von Maltzahn E, van der Merwe KJ. 1985. Improvements in the isolation of IgY from the yolks of eggs laid by immunized hens. *Immunol Investig 14*:323–327.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schaaper WMM, Lu YA, Tam JP, Meloen RH. 1989. Fine-specificity of antisera raised against a multiple antigenic peptide from foot-and-mouth disease. In: River JE, Marshall GR, eds. *Peptides: Chemistry, structure and biology. Proceedings of the 11th American Peptide Symposium*. Leiden: ESCOM.

Sellem CH, Belcour L. 1994. The in vivo use of alternate 3′-splice sites in group I introns. *Nucleic Acids Res 22*:1135–1137.

Siemeister G, Buchholz C, Hachtel W. 1990. Genes for the plastid elongation factor Tu and ribosomal protein S7 and six tRNA genes on the 73 kb DNA from *Astasia longa* that resembles the chloroplast DNA of *Euglena*. *Mol Gen Genet 220*:425–432.

Smith CWJ, Patton JG, Nadal-Ginard B. 1989. Alternative splicing in the control of gene expression. *Annu Rev Genet 23*:527–577.

Smith TF, Waterman MS. 1981. Comparison of biosequences. *Adv Appl Math 2*:482–489.

Stevenson JK, Drager RG, Copertino DW, Christopher DA, Jenkins KP, Yepiz-Plascencia G, Hallick RB. 1991. Intercistronic group III introns in polycistronic ribosomal protein operons of chloroplasts. *Mol Gen Genet 228*:183–192.

Stevenson JK, Hallick RB. 1994. The psaA operon pre-mRNA of the *Euglena gracilis* chloroplast is processed into photosystem I and II mRNAs that accumulate differentially depending on the conditions of cell growth. *Plant J 5*:247–260.

Tanaka M, Wakasugi T, Sugita M, Shinozaki K, Sugiura M. 1986. Genes for the eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): Similarity to the S10 and *spc* operons of *Escherichia coli*. *Proc Natl Acad Sci USA 83*:6030–6034.