The Plant Cell, Vol. 2, 659-671, July 1990 © 1990 American Society of Plant Physiologists

Complex RNA Maturation Pathway for a Chloroplast Ribosomal Protein Operon with an Internal tRNA Cistron

David A. Christopher[a,1] and Richard B. Hallick[a,b,2]

[a] Department of Molecular and Cellular Biology, University of Arizona, Biosciences West, Room 524, Tucson, Arizona 85721

[b] Department of Biochemistry, University of Arizona, Biosciences West, Room 524, Tucson, Arizona 85721

We have studied the expression of a large chloroplast ribosomal protein operon from *Euglena gracilis* that resembles the *Escherichia coli* S10 and *spc* ribosomal protein operons. We present evidence that 11 ribosomal protein genes, a tRNA gene, and a new locus, *orf214/orf302*, are expressed as a single transcription unit. The primary transcript also contains at least 15 group II and group III introns. Gene-specific probes for each ribosomal protein gene, *orf214/orf302*, and for *trnI* hybridized to a common pre-mRNA of an estimated size of 8.3 kilobases. This is the RNA size predicted for a full-length transcript of the entire operon after splicing of all 15 introns. Polycistronic ribosomal protein mRNAs accumulated primarily as spliced hepta-, hexa-, penta-, tetra-, tri-, and dicistronic mRNAs, which were presumed to arise by stepwise processing of the 8.3-kilobase pre-mRNA. A novel finding was the co-transcription of the *trnI* gene as an internal cistron within the ribosomal protein operon. Several combined mRNA/tRNA molecules, such as the pentacistronic *rpl5-rps8-rpl36-trnI-rps14*, were characterized. The occurrence of the *orf214/orf302* is a unique feature of the *Euglena* operon, distinguishing it from all chloroplast and prokaryotic ribosomal protein operons characterized to date. The *orf214/orf302* are not similar to any known genes but are co-transcribed with the ribosomal protein loci and encode stable RNA species of 2.4, 1.8, and 1.4 kilobases.

INTRODUCTION

Most chloroplast genes are closely spaced on the chloroplast genome and arranged in polycistronic transcription units (Ohyama et al., 1986; Shinozaki et al., 1986), which produce complex sets of overlapping mRNAs (Barkan, 1988; Westhoff and Herrmann, 1988). Many of the transcription units are organized like the comparable, evolutionarily related genes in *Escherichia coli* operons (Whitfeld and Bottomley, 1983). This is particularly true for the chloroplast-encoded ribosomal protein genes. Chloroplast (Montandon and Stutz, 1984; Fromm et al., 1986; Tanaka et al., 1986; Sijben-Muller et al., 1986; Zhou et al., 1989) and cyanelle (Evrard et al., 1989; Loffelhardt et al., 1989) genomes contain operons that resemble the *E. coli str*, S10, *spc*, and *alpha* (Cerretti et al., 1983; Lindahl and Zengel, 1986) ribosomal protein operons. In plants, 10 ribosomal protein genes are clustered in the same polarity on the chloroplast DNA as *rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-(infA)-rpl36-rps11* (Ohyama et al., 1986; Tanaka et al., 1986; Hiratsuka et al., 1989). The structure and organization of a large ribosomal protein operon on the *Euglena* chloroplast genome, as shown in Figure 1, have recently been determined (Christopher et al., 1988; Christopher and Hallick, 1989). Although the *Euglena* operon lacks *rps11*, it has two additional genes, *rpl5* and *rps14*, that are not present in the plant gene clusters but have counterparts in the *E. coli spc* operon. A *trnI* gene is located between *rpl36* and *rps14* (Nickoloff et al., 1989). In addition, *Euglena* has the unusual feature of the *orf214/orf302* locus situated in a 2.8-kb region between *rps3* and *rpl16*.

At present, despite the wealth of information on the structure of the large chloroplast ribosomal protein operons, much less is known about their overall expression. Transcripts from the spinach plastid *rps19* gene are associated with polysomes in amyloplasts (Deng and Gruissem, 1988). The *rpl2* locus appears to show an increase in its transcriptional activity relative to other nonribosomal protein genes during chloroplast development (Deng and Gruissem, 1987). The splicing of the single large group II intron of the *rpl16* locus from plant chloroplasts (Posno et al., 1986; Barkan, 1989; Zhou et al., 1989) and the group II and group III introns of the *Euglena rpl16* and *rps8* loci (Christopher and Hallick, 1989) have been studied. Groups of two to four genes within the large cluster are co-transcribed in spinach (Thomas et al., 1988; Zhou et al., 1989) and *Euglena* (Christopher and Hallick, 1989), leading to the inference that the entire S10- and *spc*-like operons in chloroplasts are cotranscribed. However, little is known

[1] Current address: Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843.

[2] To whom correspondence should be addressed.

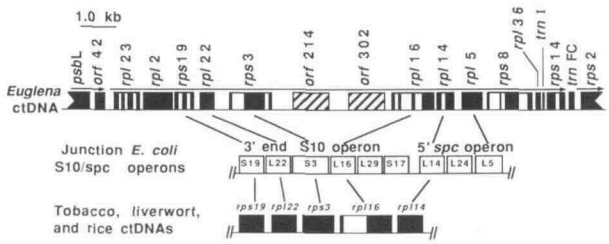

Figure 1. Maps of the *Euglena* Chloroplast Ribosomal Protein Operon, the Junction of the *E. coli* S10 and *spc* Ribosomal Protein Operons, and the Related Genes in Plant Chloroplast (ct) DNAs.

Exons are represented by black boxes, introns by open boxes (except for the *E. coli* genes). Hatched boxes represent open reading frames. In *E. coli*, the genes are identified by the ribosomal proteins that they encode. In the chloroplast DNAs, the genes follow the standard chloroplast gene nomenclature assignments (Hallick and Bottomley, 1983). The thylakoid protein genes *psbL* and *orf42* (Cushman et al., 1988a), the tRNA genes *trnFC* (Nickoloff et al., 1989), and *rps2* are included as reference genes. Diagonal lines connect evolutionarily related genes between the genomes. Arrows above the genes indicate direction of transcription.

about the mRNAs from the entire gene cluster and the actual size of their transcription unit(s). Moreover, nothing is known about the processing pathways for mRNAs derived from the genes.

Previous studies that define the mRNAs from complex chloroplast operons have been focused on abundant transcripts from genes encoding components of the photosynthetic apparatus (Berends et al., 1987; Hudson et al., 1987; Barkan, 1988; Westhoff and Herrmann, 1988). The synthesis of many of the proteins composing the photosynthetic apparatus is regulated in a light-dependent manner (Gruissem et al., 1988; Mullet, 1988), whereas the ribosomal RNAs and proteins accumulate constitutively (Deng and Gruissem, 1988). To better understand the regulation of *Euglena* chloroplast ribosomal protein gene expression, we have been studying an operon encoding 11 ribosomal protein genes. Emphasis was placed on identifying the pre-mRNAs, mature mRNAs, and the transcription units encoding them. We show that all 11 ribosomal protein genes, a tRNA gene, and a new locus, *orf214/orf302*, are cotranscribed as a single operon. Processing pathways for the pre-mRNAs and mRNAs from all the genes are proposed.

RESULTS

Abundance of *Euglena* Chloroplast *rpl16* and *psbA* Transcripts

A diagram of the overall region of the *Euglena* chloroplast DNA under study is illustrated in Figure 1. Also shown for comparison are the maps of the related genes from several plant chloroplast genomes and part of the adjacent *E. coli* S10 and *spc* ribosomal protein operons. The genes are encoded in an 11.8-kb region of the chloroplast genome (Hallick and Buetow, 1989) spanning the EcoRI restriction fragments EcoU, EcoZ, EcoJ, EcoM, and EcoA (Christopher et al., 1988; Christopher and Hallick, 1989). A first step in the analysis of the expression of the ribosomal protein genes was to determine the relative abundance of their mRNAs compared with the mRNAs from other previously studied *Euglena* chloroplast genes, such as the *psbA* gene (Hollingsworth et al., 1984). A comparison of the relative abundance of *rpl16* and *psbA* transcripts as determined by slot-blot hybridizations is shown in Figure 2. The *rpl16* transcripts are approximately 2000-fold to 3000-fold less abundant than the transcripts for *psbA*.

The Ribosomal Protein Loci Reside in a Polycistronic Transcription Unit Starting at *rpl23*

A more specific analysis of transcripts from the ribosomal protein gene clusters and the *orf214/orf302* locus was undertaken. The *orf214/orf302* locus consists of two open reading frames of 214 and 302 amino acids, which reside in the *rps3-rpl16* intercistronic region. First, to determine whether the 5′-upstream genes *psbL* and *orf42* (Figure 1) encoding photosystem II polypeptides are cotranscribed with the ribosomal protein loci *rpl23*, *rpl2*, *rps19*, *rpl22*, and *rps3*, RNA blot hybridization and primer extension analyses were performed. Five radiolabeled ribosomal protein gene-specific RNA probes were used in RNA gel blot

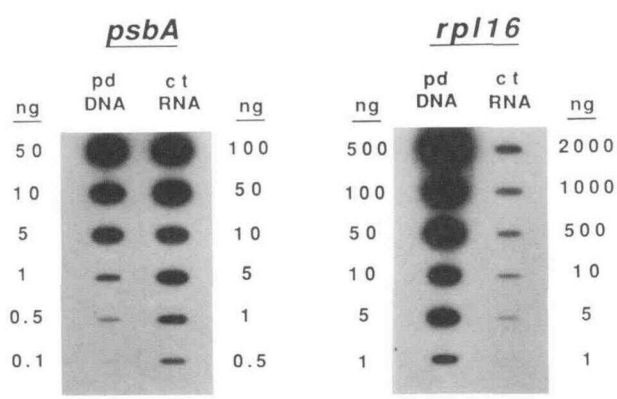

Figure 2. Relative Abundance of *Euglena* Chloroplast *psbA* and *rpl16* Transcripts.

Slot blots contained serial dilutions (in nanograms) of either chloroplast RNA (ctRNA) or gene-specific plasmid DNA (pd DNA, from plasmid clones described below) and are labeled to the side of each sample. Radiolabeled RNA probes for *psbA* were derived from plasmid pEZC514.9 (in Karabin et al., 1984) and for *rpl16* from plasmid pEZC541.2.4-4 (Table 2).

hybridization experiments, as illustrated in Figure 3A. Transcripts of 0.3, 0.7, 0.8, 1.0, 1.1, 1.3, 1.5, 1.9, 2.1, 2.4, 3.1, 4.5, 5.0, 5.5, 7.8, and 8.3 kb and additional minor bands were detected. In a control experiment, transcripts of 6.5 kb, 3.3 kb, and 1.1 kb were detected with the *psbL-orf42* probe from the upstream operon. The 6.5-kb and 3.3-kb RNAs are unique in size and not observed with the ribosomal protein gene probes (Figure 3A, lanes B through F).

Because several large RNAs were detected on the RNA gel blots by using the *rpl23* probe (Figure 3A), it was important to determine whether these RNAs had a common 5′ end and whether they overlapped with *psbL-orf42*, which is separated from *rpl23* by a 148-bp intercistronic spacer. The 5′ end of *rpl23* was mapped by primer extension cDNA sequencing of chloroplast RNA, as shown in Figure 4. A major stop site was observed just after the sequence 5′-AACAA-3′, which resides 25 nucleotides (nt) upstream from the initiator methionine-ATG of *rpl23* and 124 nt downstream from *orf42*. On overexposed gels, a minor stop site could also be detected at −55 from the ATG (data not shown). No transcripts overlapping the *orf42* and *rpl23* loci could be detected in either the RNA gel blot hybridization or primer extension experiments. These results are consistent with the previous designations (Cushman et al., 1988b) for the mRNAs detected with the *psbL-orf42* probe: The 6.5 kb is the *psaA-psaB-psbE-psbF-psbL-orf42*, the 3.3 kb is the *psaB-psbE-psbF-psbL-orf42*, and the 1.1 kb is the *psbE-psbF-psbL-orf42*.

Polycistronic pre-mRNAs common to all five ribosomal protein cistrons are clearly evident. The 3.1-kb to 8.3-kb RNAs were detected with each ribosomal protein gene probe (Figure 3A, lanes B through F). The 5.0-kb to 8.3-kb transcripts are much larger than the region encoding the five ribosomal protein genes and are presumed to be large precursor mRNAs, including transcripts of loci distal to *rps3*. Figure 5 shows that transcripts of similar size were also detected with the probes specific to *orf214* and *orf302*. A detailed analysis of these mRNAs is described below.

Polycistronic Chloroplast Ribosomal Protein mRNAs Accumulate Primarily as Penta-, Tetra-, Tri-, and Dicistronic Species

The identities of the transcripts and the proposed processing pathway deduced from the RNA gel blot hybridization experiments (Figure 3A) are shown in Figure 3B. A 4.5-kb RNA was detected with five different ribosomal protein gene probes (Figure 3A, lanes B through F) and the *orf214* and *orf302* probes (Figure 5). This is interpreted as a pentacistronic *rpl23-rpl2-rps19-rpl22-rps3* RNA plus the mature 1.4-kb transcript from the *orf214/orf302* loci described below and shown in Figure 5. The 3.1-kb RNA (Figure 3A) was only detected with probes for *rpl23*

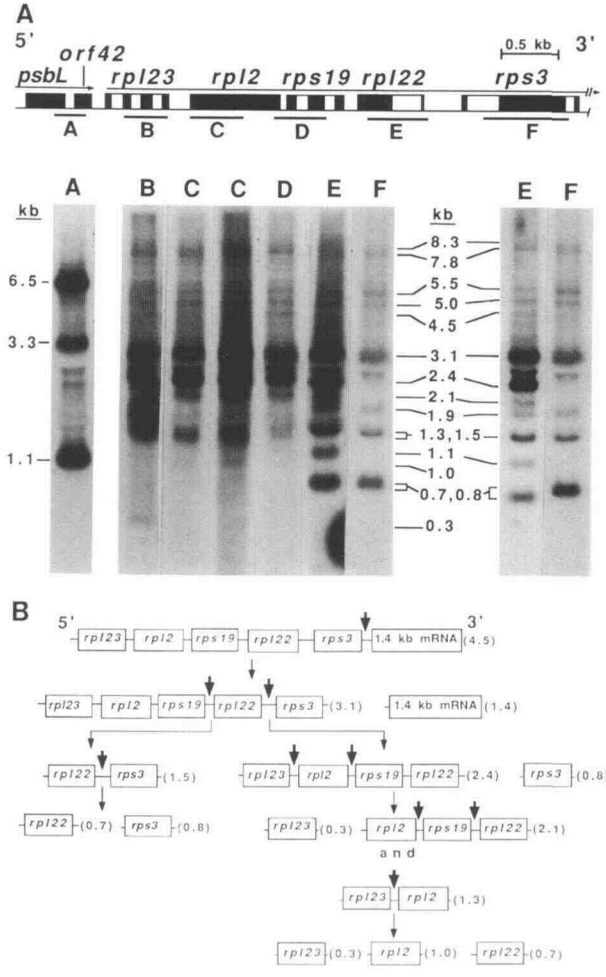

Figure 3. RNA Blot Hybridizations of *Euglena* Chloroplast mRNAs from Thylakoid Protein Genes (*psbL*, *orf42*) and Ribosomal Protein Genes (*rpl23*, *rpl2*, *rps19*, *rpl22*, *rps3*).

(A) A gene map of the region under study is presented. Long arrows above the genes indicate direction of transcription. Exons are represented by black boxes, introns by open boxes. The specific radiolabeled RNA probes used in the RNA gel blot hybridizations are marked as black lines under the genes and are identified by capital letters (probes listed in Table 2). Each lane represents a separate RNA blot (12 μg) and is labeled with the letter corresponding to the specific probe utilized in the hybridization. The pair of blots in the extreme right panels, E and F, contained RNAs from longer electrophoresis runs. The signals corresponding to mRNAs detected in the hybridizations are labeled by estimated size in kilobases.

(B) The proposed mRNA processing pathway as deduced from the RNA gel blot hybridizations in **(A)**. Transcripts are symbolized by open boxes connected with thin lines. Individual cistrons are designated by their gene names. Sizes (≤4.5 kb) are indicated in parentheses to the right of each transcript and correspond to the labeled signals in **(A)**. Thick arrowheads represent suggested intercistronic processing sites.

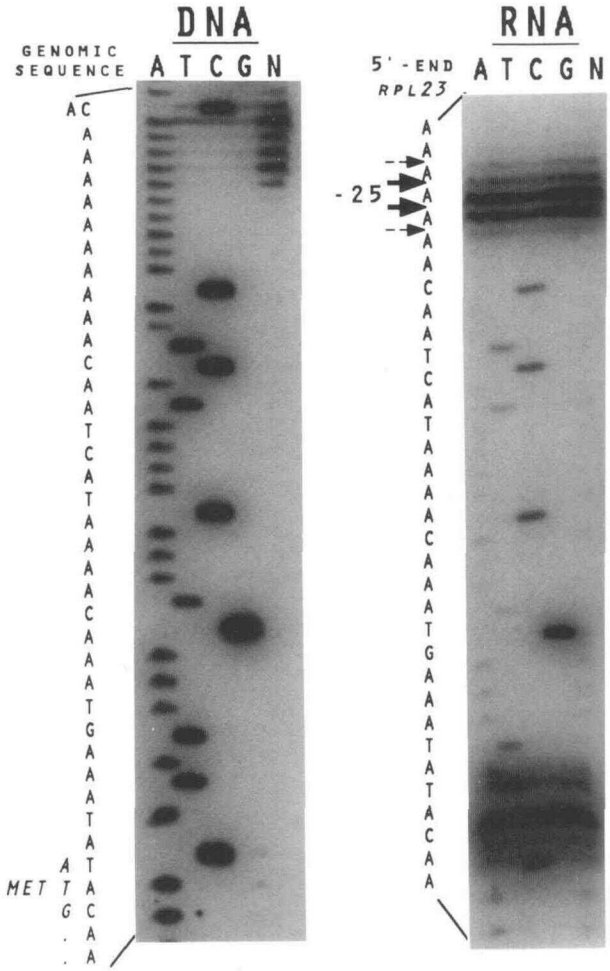

Figure 4. Primer Extension Mapping of the 5′ Terminus of the Transcripts Encoding *rpl23*.

The lanes labeled RNA (A,T,C,G,N) in the right panel contain the cDNA sequence generated by reverse transcription of *Euglena* chloroplast RNA using a primer complementary to *rpl23* exon 1. The cDNA sequence of the *rpl23* leader is listed to the left of the lanes from which it is read. The sequence stops at position −25 nucleotides upstream from the translational start of *rpl23* (denoted by arrowheads). The left panel contains the chloroplast genomic DNA sequence for the *rpl23* leader that is produced using the same primer as for the transcript analysis.

through *rps3*. The 3.1-kb RNA is the size predicted for the fully spliced pentacistronic *rpl23-rpl2-rps19-rpl22-rps3* RNA. Intercistronic cleavage between *rps3* and the 1.4-kb RNA encoding *orf214* (Figure 5) would result in two products: the 3.1-kb pentacistronic *rpl23-rpl2-rps19-rpl22-rps3* and the 1.4-kb RNA specific to the *orf214-orf302* locus. The 2.4-kb RNA is interpreted as the tetracistronic *rpl23-rpl2-rps19-rpl22* RNA. It was not detected with the *rps3*-specific probe. The tetracistronic 2.4-kb RNA is interpreted as resulting from the endonucleolytic cleavage of the monocistronic *rps3* (approximately 0.7 kb to 0.8 kb in size) from the pentacistronic 3.1-kb mRNA (Figure 3B). The 2.1-kb RNA was detected only by the use of probes from *rpl2*, *rps19*, and *rpl22*. It is the size predicted for the fully spliced tricistronic *rpl2-rps19-rpl22* mRNA. The removal of the 0.3-kb monocistronic *rpl23* from the tetracistronic 2.4-kb RNA would yield the tricistronic 2.1-kb RNA species. The 1.9-kb RNA was only detected by using probes specific to *rpl22* and *rps3*. It equaled the size of the dicistronic pre-*rpl22-rps3*, whereas the 1.5-kb RNA is inferred to be the dicistronic mature *rpl22-rps3* mRNA.

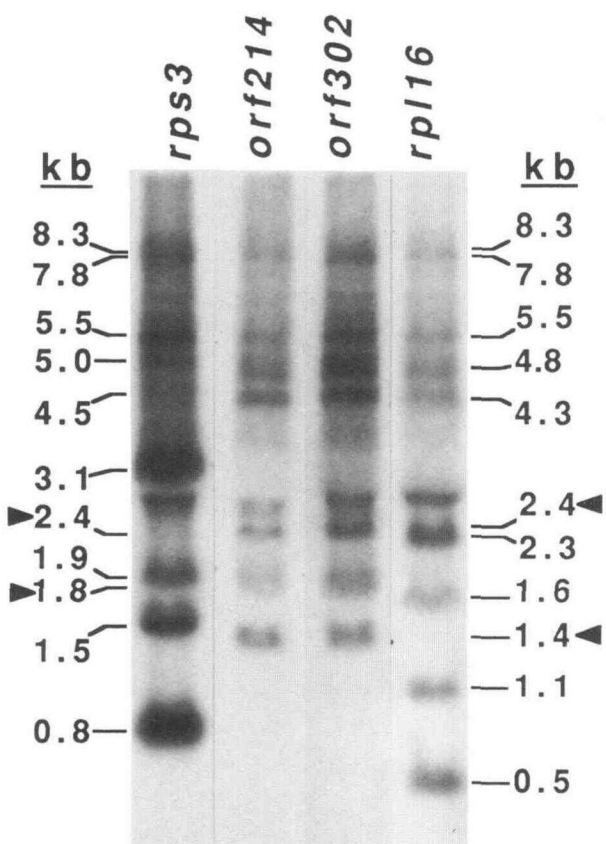

Figure 5. RNA Gel Blot Hybridization of the *Euglena* Chloroplast mRNAs from the *rps3* and *rpl16* Loci and the *orf214* and *orf302* Loci Present in the 2.8-kb *rps3-rpl16* Intercistronic Region (Figure 1).

Radiolabeled RNA probes were synthesized from plasmid clones containing the gene-specific inserts. The lanes are labeled with the gene probes used in the hybridization. The signals corresponding to mRNAs detected in the hybridization experiments are indicated by size in kilobases. Arrowheads designate unique mRNAs from the *orf214* and *orf302* loci.

The 1.3-kb RNA was detected only with the probes containing sequences from *rpl23* and *rpl2* mRNA and is of the size predicted for the fully spliced dicistronic *rpl23-rpl2* mRNA. It also appeared in lane D (*rpl2-rps19* probe). This hybridization was probably due to the *rpl2* sequences present on this probe. The 1.1-kb and 0.7-kb RNAs were only observed by using the *rpl22* probes. The 1.1-kb RNA equaled the predicted size of monocistronic pre-*rpl22*, whereas the 0.7-kb RNA corresponded in size to mature monocistronic *rpl22* mRNA. A faint signal of 1.0-kb corresponding to the predicted size of monocistronic *rpl2* was detected in lane C using the *rpl2*-specific probe (Figure 3A) and was more discernible on a longer exposure of the blot.

The monocistronic 0.7-kb *rpl22* and 0.8-kb *rps3* RNAs can be resolved from each other via longer electrophoresis (Figure 3A). These RNAs appear to be quite abundant. They accumulate to higher levels compared with the faint bands interpreted to be the monocistronic *rpl23* (0.3 kb) and *rpl2* (1.0 kb). In contrast to the distinct monocistronic *rpl22* and *rps3* RNA species, no monocistronic-sized RNA species corresponding to the *rps19* (0.3 kb) was detected (Figure 3A, lane D).

The *orf214/orf302* Loci Are Cotranscribed with Flanking Ribosomal Protein Genes

The close linkage of the ribosomal protein genes *rps3* and *rpl16* is evolutionarily conserved among prokaryotes and plant chloroplast genomes (Figure 1). The occurrence of the *orf214/orf302* loci in the *Euglena rps3-rpl16* intercistronic region was unexpected and is highly novel. To determine whether the *orf214/orf302* loci are transcribed and to compare the RNA transcripts with those from the ribosomal protein loci, RNA gel blot analysis was undertaken utilizing probes specific to *orf214*, *orf302*, *rps3*, and *rpl16* (Figure 5). The 2.8-kb *rps3-rpl16* intercistronic region is clearly transcribed. Probes from *orf214* and *orf302* hybridized to RNAs of the same polarity as the ribosomal protein genes. No transcripts were detected using probes of the opposite polarity (data not shown). Hence, the DNA-coding strand for the *orf214/orf302* locus is the same strand as for the ribosomal protein genes. The 2.4-kb, 1.8-kb, and 1.4-kb RNA species appear to be unique to the *orf214* and *orf302* loci. These RNAs were detected with either the *orf214* or *orf302* probes but not with the probes from *rps3* and *rpl16* (Figure 5). They were inferred to originate from post-transcriptional processing events of the larger 4.3-kb to 8.3-kb RNAs; however, transcription from other promotors cannot be excluded. Transcription initiation was not observed in the *orf214/orf302* locus nor in the 5′- and 3′-flanking regions (G. Yepiz-Plascencia and R.B. Hallick, manuscript in preparation).

The 8.3-kb, 7.8-kb, and 5.5-kb RNA species detected with *orf214* and *orf302* probes are larger than the 2.8-kb *rps3-rpl16* intercistronic region and are the same size as the transcripts detected with the *rps3* and *rpl16* probes. The presence of these common RNAs is evidence for cotranscription of the ribosomal protein genes with the *orf214/orf302* locus. Additional common RNA species of 4.3 kb, 4.5 kb, 4.8 kb, and 5.0 kb were observed with probes from *orf214*, *orf302*, *rpl16*, and *rps3*. These transcripts were interpreted to be polycistronic precursor mRNAs encoding both ribosomal protein genes and the *orf214* and *orf302* loci.

The *rpl16*, *rpl14*, *rpl5*, *rps8*, *rpl36*, and *rps14* Loci, Including tRNA-Ile, Are Cotranscribed with the Upstream Five Ribosomal Protein Loci

The *rpl16*-specific probe hybridized with pre-mRNAs of the same size as detected by using the *rps3* probe (Figure 5), as expected if these genes are within the same operon. To further examine this possibility, RNA gel blot hybridization experiments were conducted using gene-specific probes from *rpl16*, *rpl14*, *rpl5*, *rps8*, *rpl36*, *trnI*, and *rps14*, as shown in Figure 6A. Transcripts of 0.4, 0.5, 0.6, 0.7, 1.0, 1.1, 1.2, 1.6, 1.8, 2.1, 2.3, 2.7, 4.3, 4.8, 5.5, 7.8, and 8.3 kb were detected. The RNA species 4.3 kb through 8.3 kb were detected with all seven gene-specific probes (Figure 6A, lanes K to P, and *trnI*). These RNAs are interpreted to be common polycistronic precursor transcripts spanning the *rpl16-rps14* region and the upstream regions up to and including *rpl23*. The processing of these common transcripts is believed to yield the smaller RNAs (0.07 kb to 2.7 kb), as defined below and illustrated in Figure 6B.

The 2.7-kb RNA hybridized to probes from *rpl16* through *rps14* (Figure 6A). The 2.7-kb RNA is the expected size of the spliced exons plus intergenic regions for the *rpl16* through *rps14* loci. Unspliced RNAs were not detected in primer extension sequencing experiments conducted across the splice junctions of these transcripts (Christopher and Hallick, 1989). The 2.7-kb RNA is, therefore, interpreted to be a fully spliced heptacistronic transcript encoding *rpl16-rpl14-rpl5-rps8-rpl36-trnI-rps14*. The 2.3-kb RNA hybridized with individual probes specific for *rpl14*, *rpl5*, *rps8*, and *trnI* and more general probes from *rpl36-trnI-rps14* and *rps14-trnF* (Figure 6A, lanes L to P). The 2.3-kb RNA did not appear on RNA gel blots hybridized with probes from *rpl16* or *trnC* and *rps2*. The 2.3-kb RNA is interpreted as a polycistronic *rpl14-rpl5-rps8-rpl36-trnI-rps14* mRNA.

The 2.1-kb RNA hybridized with the *rpl16*, *rpl14*, *rpl5*, and *rps8* probes (Figure 6A, lanes K to N) but not with probes from the region 3′ to *rps8* (lanes O through R). The 2.1-kb RNA was thus inferred to be the tetracistronic *rpl16-rpl14-rpl5-rps8* mRNA. The 1.8-kb RNA was detected with probes from *rpl5* through *rps14*. It did not hybridize to probes from *rpl16* or *rpl14* and is suggested to be the pentacistronic *rpl5-rps8-rpl36-trnI-rps14*. The

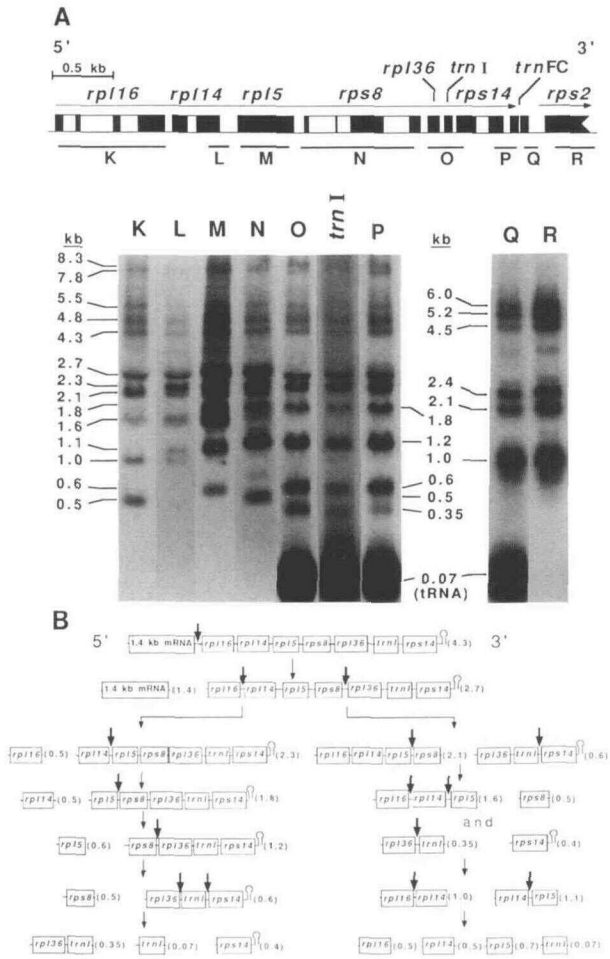

Figure 6. RNA Gel Blot Hybridizations of *Euglena* Chloroplast RNAs from Ribosomal Protein and tRNA Genes.

(A) From the gene map, the exons are symbolized by black boxes and the introns by open boxes. Long arrows above the genes indicate direction of transcription. The probes used in the hybridizations are represented by a black line under each gene(s) and are identified by capital letters (probes listed in Table 2). Each lane represents a separate RNA blot and is labeled with the letter corresponding to the specific probe utilized in the hybridization. The lane labeled *trnI* refers to a synthetic *trnI*-specific oligomer probe of 55 nt used on that particular blot. The signals corresponding to RNAs detected in the hybridizations are labeled by size in kilobases.

(B) The proposed mRNA processing pathway as deduced from the hybridizations in **(A)**. Transcripts are symbolized by open boxes connected with thin lines. Individual cistrons are designated by the genes from which they arose. Sizes (≤4.3 kb) are indicated in parentheses to the right of each transcript and correspond to the labeled signals in **(A)**. Thick arrowheads represent suggested intercistronic processing sites.

1.6-kb RNA was detected only when probes from *rpl16*, *rpl14*, and *rpl5* were used. A faint 1.6-kb RNA appeared in lane N (*rps8* probe), probably because this probe spans a few nucleotides at the 3′ end of *rpl5*. In addition, the 1.6-kb RNA was too small to also encode *rps8* and was inferred to be tricistronic *rpl16-rpl14-rpl5*. The 1.2-kb RNA hybridized to probes from *rps8* through *rps14*, including a *trnI*-specific 55-nt synthetic oligomer, and was defined as a tetracistronic *rps8-rpl36-trnI-rps14*. The 1.1-kb and 1.0-kb RNAs (lanes K, L, and M) are predicted to be dicistronic *rpl14-rpl5* and *rpl16-rpl14*, respectively. The 0.6-kb RNA common to lanes O to P hybridized with probes from *rpl36*, *trnI*, and *rps14*. It is the predicted size of a tricistronic RNA from these genes. The 0.35-kb band is interpreted as two comigrating mRNA species, dicistronic *rpl36-trnI* and monocistronic *rps14*.

Monocistronic RNA species for *rpl16* (0.5 kb), *rpl14* (0.5 kb), *rpl5* (0.6 kb), *rps8* (0.5 kb), and *trnI* (70 nt) were observed as distinct bands that corresponded to the sizes of the fully spliced exons plus a short 5′ leader. The monocistronic *rpl36* was not observed. The *trnI*-specific oligonucleotide probe hybridized to the monocistronic *tRNA-Ile-CAU* and the 0.35-, 0.6-, 1.2-, 1.8-, 2.3-, 2.7-, 4.3-, 4.8-, 5.5-, 7.8-, and 8.3-kb RNA species (Figure 6A, lane *trnI*). The *trnI* gene is, therefore, an internal cistron within a large polycistronic operon encoding ribosomal protein genes.

The proposed processing events for these RNAs and the interpretation of the identities of the RNA species (sizes 0.07 kb to 4.3 kb) from the *rpl16* through *rps14* region are presented as a processing pathway in Figure 6B. In this proposed pathway, the 4.3-kb RNA encodes sequences mapping between *orf214* and *rps14*, inclusive (Figure 1). The 4.3-kb RNA is predicted to be processed upstream from *rpl16* to yield the 1.4-kb transcript from *orf214/orf302* (shown in Figure 5) and the 2.7-kb heptacistronic *rpl16-rpl14-rpl5-rps8-rpl36-trnI-rps14*. Two alternate cleavage events are then proposed to occur in the heptacistronic RNA (Figure 6B). One in the *rpl16-rpl14* intergenic spacer produces the 0.5-kb monocistronic *rpl16* and the 2.3-kb *rpl14-rpl5-rps8-rpl36-trnI-rps14*. Further cleavage events are envisioned to occur progressively one cistron at a time from the 5′ end. The other proposed pathway starts with an endonucleolytic cleavage between *rps8* and *rpl36*. The two RNAs produced, a 2.1-kb tetracistronic *rpl16-rpl14-rpl5-rps8* and a 0.6-kb tricistronic *rpl36-trnI-rps14*, are further processed to tri-, di-, and monocistronic RNAs as defined above.

The 5.5-kb, 7.8-kb, and 8.3-kb RNAs are the same size as, and comigrate with, the largest RNAs from the upstream ribosomal protein genes (Figure 3A) and the *orf214/orf302* locus (Figure 5). These RNAs were not detected with probes from the region 3′ distal to *rps14*. Results with two such probes, one spanning the *trnC* plus *trnC-rps2* intergenic spacer (lane Q) and the other specific for

rps2 (lane R) are presented in Figure 6A. Signals corresponding to the 7.8-kb and 8.3-kb RNAs are absent from these blots. Therefore, the *rps2*, *trnC*, and probably *trnF* are not likely to be cotranscribed with the upstream ribosomal protein gene operon. We conclude that at least 12 genes from a single ribosomal protein operon are transcribed starting at *rpl23* and terminating after *rps14*. The location of a possible termination site 3′ to *rps14* is further supported by S1 nuclease analysis of the *rps14-trnF* spacer region.

The 3′ End of *rps14*

Because probes for genes downstream from *rps14* (*trnC* and *rps2*) did not hybridize with the transcripts from the large ribosomal protein gene operon, the transcription unit most likely terminates after *rps14*. To determine the 3′ end of the transcripts detected with probes upstream from, and including, *rps14*, S1 nuclease protection experiments were performed. A fully protected fragment was not obtained with a probe spanning the intercistronic region from *rps14* exon 2 to *trnF*, as illustrated in Figure 7. The protected fragments were for *trnF* (56 nt to 57 nt) and for *rps14* exon 2 plus 30 nt of its 3′-flanking sequences (123 nt to 124 nt), which includes a stem-loop region (Nickoloff et al., 1989). Hence, overlapping transcription between *rps14* and *trnF* was not observed. The 3′ ends of the large polycistronic transcripts detected using *rps14* probes on the RNA gel blots mapped to the region between the 3′ end of *rps14* and the *trnF*. The *trnF-trnC* genes can be transcribed as a single dicistronic operon, separate from the *rps14* gene, in in vitro chloroplast transcription extracts (Greenberg and Hallick, 1986). The *trnFC* and distal genes are, most likely, not part of the upstream operon that begins with *rpl23* and ends with *rps14*. We cannot formally rule out the possibility that *trnFC* might be cotranscribed with the upstream genes, but rapidly processed.

Additional transcription units are believed to start at the *trnFC* and *rps2* loci. An *rps2*-specific probe (Figure 6A, lane R) hybridized to several mRNAs, 6.0, 5.2, 4.5, 2.4, 2.1, and 1.0 kb. The *rps2* gene appears to be cotranscribed with downstream genes (*atpI-atpH-atpF-atpA*) from the chloroplast ATP synthase operon (R.G. Drager and R.B. Hallick, unpublished data). The 1.0-kb mRNA present in lanes Q and R (Figure 6A) is the size predicted for a spliced monocistronic *rps2* mRNA. It hybridized with the *rps2*-specific probe and a probe (lane Q) containing sequences from *trnC* and the intergenic spacer between *trnC* and *rps2*. Probe Q lacks *rps2* coding sequences, yet it still hybridized to the same RNAs as the *rps2* probe. This hybridization could be due to the presence of the *trnC-rps2* intergenic spacer, which might actually include the 5′ leader for *rps2*. The extent of the *rps2* leader is currently being determined.

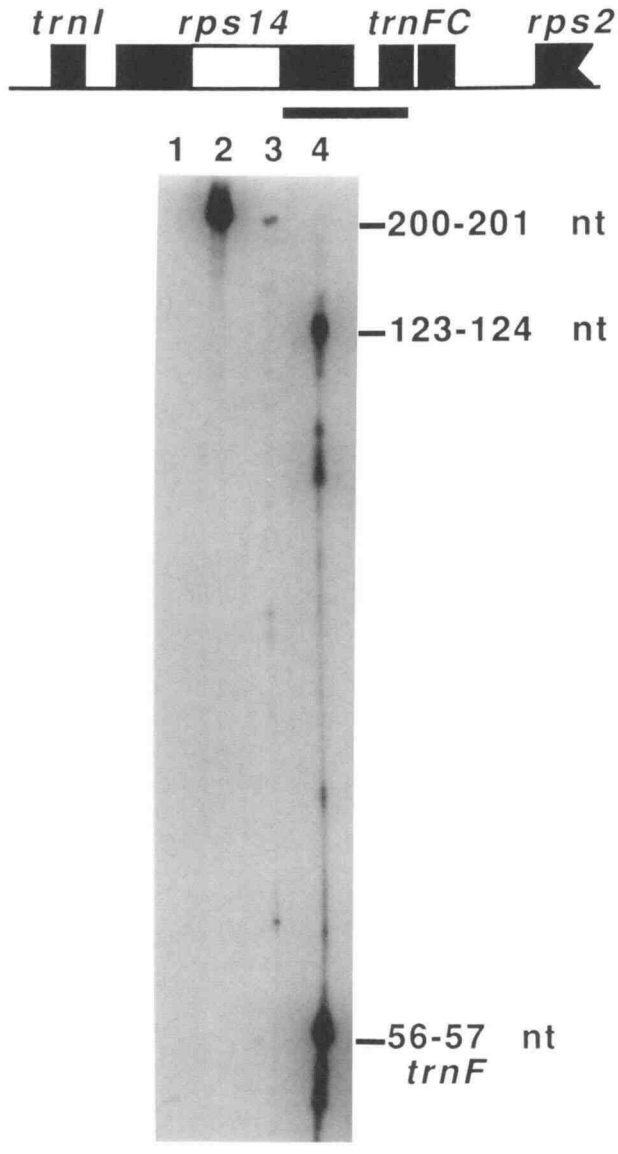

Figure 7. S1 Nuclease Mapping of the 3′ End of *rps14* in the *rps14-trnF* Intercistronic Spacer.

In the gene map, exons are represented by black boxes and introns by open boxes. The specific radiolabeled RNA probe utilized in the S1 nuclease digestions (201 nt, from the plasmid pEZC800.7-15, Table 2) is symbolized by a black line under the genes and is shown in lane 2. It spans 93 nt of *rps14* exon 2,52 nt of intergenic spacer, and 56 nt of *trnF*. Lane 1 shows the probe annealed to yeast tRNA and digested with S1 nuclease. Lane 3 shows the probe annealed to 200 ng of plasmid pEZC800.7-15 and digested with S1 nuclease. Lane 4 shows the probe annealed to 6 μg of purified chloroplast RNA and digested with S1 nuclease.

DISCUSSION

A Novel *Euglena* Chloroplast Ribosomal Protein Operon

The expression of a large *Euglena* chloroplast operon encoding 11 ribosomal protein genes, the *orf214/302* locus, and a tRNA has been analyzed. Emphasis was placed on characterizing the RNAs and the RNA-processing pathways for these genes. We have demonstrated cotranscription of 11 ribosomal protein genes and of an internal tRNA cistron in a polycistronic mRNA. Information available on the expression of similar ribosomal protein gene clusters present in the tobacco (Tanaka et al., 1986), liverwort (Ohyama et al., 1986), and rice (Hiratsuka et al., 1989) chloroplast genomes has previously been confined to a few genes within the operon. For example, run-on transcription assays of lysed spinach chloroplasts yields labeled *rpl2* mRNAs. The transcriptional activity of *rpl2* increases relative to other genes (*psbA*, *rbcL*, and *atpB/E*) during chloroplast development (Deng and Gruissem, 1987). In spinach chloroplasts, cotranscription of the four completely characterized genes, *rpl22*, *rps3*, *rpl16*, and *rpl14*, has been observed (Zhou et al., 1989). These cistrons are proposed to be cotranscribed with the upstream *rpl23*, *rpl2*, and *rps19* loci (Thomas et al., 1988). However, additional downstream loci such as *rps8*, *rpl36*, and *rps11* must be characterized in spinach before a complete model of their expression can be formulated. Based on the results with the related *Euglena* genes presented here, it seems reasonable to suggest that cotranscription of the higher plant genes occurs.

The organization of the 11 ribosomal protein genes in a single transcription unit is a remarkable feature that may reflect unique aspects of *Euglena* chloroplast gene expression. Being composed of 11 ribosomal protein genes and a tRNA gene, including the *orf214/orf302* locus and a minimum of 15 introns, this is one of the most complex chloroplast operons identified to date. In *E. coli*, most of the 53 ribosomal protein genes are organized in polycistronic operons (Nomura et al., 1977; Lindahl and Zengel, 1986). The simultaneous expression of the gene clusters can be regulated at a few common control sites. The organization of the *Euglena* genes resembles the two *E. coli* ribosomal protein operons S10 and *spc*. Distinct transcription initiation and termination sites separate the *E. coli* operons (Lindahl and Zengel, 1986); however, a certain amount of partial transcriptional read-through occurs from the S10 into the *spc* (Mattheakis and Nomura, 1988) and from the *spc* into the alpha operons (Lindahl and Zengel, 1986). Hence, there is precedence for overlapping transcription between the *E. coli* operons. Evidently, the control sites that delimit the operon boundaries in *E. coli* have not, however, been formed and conserved during the evolution of the *Euglena* chloroplast operon.

Cotranscription of a tRNA with Genes Coding for Protein

A highly novel feature of the gene expression in this region is the cotranscription of the *trnI* gene as an internal cistron within the ribosomal protein operon. The *trnI* is framed, 5′- and 3′-, by ribosomal protein loci with which it is cotranscribed. Several stable polycistronic RNAs accumulate that also contain the tRNA-Ile. The tRNA is presumably released by endonucleolytic cleavage on either side of the tRNA in the precursor mRNA/tRNA molecule. The differential accumulation of the tRNA relative to the mRNAs could simply be due to the increased stability of the tRNA over the mRNAs. One of the mechanisms for the differential expression of chloroplast genes has been attributed to differences in post-transcriptional RNA stability and processing (Stern and Gruissem, 1987). The possibility cannot be ruled out that *trnI* is transcribed as an internal monocistronic locus in addition to being cotranscribed with the ribosomal protein loci. This seems less plausible, however, because the short 28-bp *rpl36-trnI* intercistronic spacer (Christopher and Hallick, 1989; Nickoloff et al., 1989) does not contain the 5′-flanking sequences of the prokaryotic-like promoter elements previously described for chloroplast tRNA genes (Gruissem and Zurawski, 1985; Gruissem et al., 1986b; Hanley-Bowdoin and Chua, 1987). Eukaryotic-like promotor elements internal to the tRNA could exist (Gruissem et al., 1986a), but this requires confirmation using chloroplast extracts to transcribe in vitro mutagenized tRNA loci. The RNA gel blot hybridization experiments described here do not provide data on the transcriptional activities of the genes relative to other chloroplast genes.

In plants, the organization of protein coding loci in transcription units along the chloroplast chromosome is punctuated by tRNA genes that are nonclustered (Ohyama et al., 1986; Shinozaki et al., 1986; Hiratsuka et al., 1989) and apparently transcribed independently from flanking protein genes. In *Euglena* chloroplasts, the tRNA genes lie between protein and rRNA coding regions and are encoded as polycistronic clusters of two to six different cistrons (Hallick and Buetow, 1989). The *trnI*-CAU is an exception to this general feature. None of the known plant or *Euglena* chloroplast tRNA genes has been documented to be cotranscribed with protein coding loci. In the mammalian mitochondrial genome, the tRNA genes are cotranscribed with the genes encoding protein and rRNA from a single promoter (Clayton, 1984). The 5′ and 3′ ends of tRNAs appear to serve as specific processing sites within the large pre-mRNA. The tRNA genes of *E. coli* are clustered, some of which are in operons transcribed from a single promoter. Mixed tRNA/mRNA and tRNA/rRNA operons do exist in *E. coli*. One example is the *tufB* operon, which encodes a cluster of four tRNA loci and the protein elongation factor, EF-TuB (Lindahl and Zengel, 1986). The *Euglena* transcripts containing tRNA-Ile are, therefore, the

first examples in chloroplasts of mixed tRNA/mRNA molecules in which a tRNA is embedded within a polycistronic protein coding mRNA.

Codon Bias for tRNA-I-CAU in mRNAs for Constitutive Versus Light-Induced Genes

One biological basis for including a *trnI*-CAU within the ribosomal protein operon would be the increased utilization of this tRNA during ribosomal protein mRNA translation. There are two tRNAs for isoleucine in *Euglena* chloroplasts. Based on wobble codon:anticodon recognition in chloroplasts, the tRNA-I-CAU present in the operon described here is believed to read the codon ATA (Kashdan and Dudock, 1982). The tRNA-I-GAU reads the other two isoleucine codons, ATC and ATT. In Table 1, the isoleucine codon usage is compared between two types of chloroplast genes, constitutive and light dependent. The chloroplast genes for ribosomal proteins (i.e., *rpl23*, *rpl16*) and RNA polymerase subunits (i.e., *rpoB*, *rpoC*) are considered to be constitutively expressed because their genes are expressed and their products are constitutively present among the various plastid developmental stages (Thompson and Whatley, 1980; Deng and Gruissem, 1988). In contrast, the appearance of polypeptides composing the photosynthetic apparatus is a light-induced process (Mullet, 1988). From among all 27 genes analyzed, isoleucine codons ending in T or C are generally preferred overall; however, there is a distinct minimum twofold preference for the use of the ATA codon and, hence, the tRNA-I-CAU in the 18 constitutively expressed genes. This preference is evident as a marked codon bias for Ile-ATA in genes for constitutively expressed versus light-induced chloroplast proteins. The *trnI*-CAU is the only tRNA for which this codon bias-gene relationship consistently applies (Christopher et al., 1988). The presence of the *trnI*-CAU within the ribosomal protein operon may make this particular operon indispensable for *Euglena* chloroplasts because this is the only copy of a gene that apparently reads a codon that occurs in higher proportion in the constitutively expressed genes versus light-dependent genes. Further experimental analysis is required to determine whether the cotranscription of the *trnI*-CAU with the ribosomal protein loci is a mechanism of coregulating the supply of this tRNA over other tRNAs and, in turn, regulates the translational efficiency of mRNAs differing in codon usage (Dix and Thompson, 1989). Certainly, it may be a means of fine-tuning translation within an array of other more prominent translational regulatory mechanisms (Klein et al., 1988).

Table 1. Comparison of Isoleucine Codon Usage between Constitutive and Light-Dependent *Euglena* Chloroplast Genes

	Isoleucine Codon Usage		
Genes	ATT/C	ATA	ATT/C:ATA
Constitutive			
rpl23, rpl2, rps19, rpl22, rps3	49	29	1.7:1
rpl16, rpl14, rpl5, rps8, rpl36, rps14	41	25	1.6:1
rpoB	61	37	1.6:1
rpoC1/C2	88	52	1.7:1
tufA, rps12, rps7	32	33	1.0:1
rps2	14	11	1.3:1
Light dependent			
rbcL	18	2	9.0:1
psbA	26	2	13.0:1
psaA, psaB	74	31	2.4:1
psbE, psbF, psbL, orf42	16	3	5.3:1
atpH	25	7	3.6:1

The isoleucine codons ATT and ATC are combined and their frequency of occurrence is compared with codon ATA. This comparison is presented as a ratio in the last column. References: *rpl23* to *rps3* in Christopher et al. (1988); *rpl16* to *rpl36* in Christopher and Hallick (1989); *rps14* in Nickoloff et al. (1989); *rpoB* in Yepiz-Plascencia et al. (1990); *rpoC1/C2* from C.A. Radebaugh and R.B. Hallick, manuscript in preparation; *rps2* and *atpH* from R.G. Drager and R.B. Hallick, manuscript in preparation; *tufA*, *rps12*, and *rps7* in Montandon and Stutz (1983, 1984); *psbA* in Karabin et al. (1984); *rbcL* in Gingrich and Hallick (1985); *psaA*, *psaB*, and *psbE* to *orf42* in Cushman et al. (1988a, 1988b).

The *orf214; shorf320* Locus

The interruption of the ribosomal protein transcription unit by the *orf214/orf302* and its cotranscription with the ribosomal protein loci are remarkable features of the *Euglena* chloroplast DNA, distinguishing it from among all chloroplast genomes and partial genomes characterized to date. Likewise, this region in *Euglena* has no counterpart in the analogous ribosomal protein operons from prokaryotic genomes (Lindahl and Zengel, 1986). In the chloroplast and prokaryotic genomes, the *rps3* and *rpl16* genes are contiguous (Figure 1). The identification of a new locus in such a highly conserved chloroplast ribosomal protein operon is noteworthy. A 1.54-kb mRNA would be needed to encode the 516 amino acids predicted in the *orf214/orf302* locus. The 1.4-kb mRNA is most coincident with this predicted size. From the DNA sequence of the region (D.A. Christopher and R.B. Hallick, unpublished data), structures resembling group II intron domains five and six have been found in this region. The possibility of post-transcriptional processing and splicing of pre-mRNAs is suggested to give rise to the 1.4-kb mRNA. The polypeptides predicted from the *orf214/orf302* locus are not highly similar to any other known genes. Hence, the genetic function of this unique region is unknown at present. Experiments designed to examine the *orf214/orf302* further, including cDNA sequence analysis, are underway.

RNA Processing Pathways: Ribosomal Protein mRNAs and Pre-mRNAs Accumulate Primarily as Spliced Polycistronic Transcripts That Are Further Processed

The ribosomal protein operon gives rise to a complex pattern of overlapping mRNAs. A series of RNA processing pathways are proposed beginning with the 8.3-kb RNA, the largest common transcript detected with all the gene probes, and ending with the various monocistronic RNAs. The working hypothesis includes larger RNAs being processed and converted into smaller RNAs by endonucleolytic cleavage events that appear to occur in the intercistronic spacers. Because 5′ and 3′ ends of the numerous RNA species were not mapped, other interpretations of these data are also possible. Chloroplast extracts contain many processing activities (Gruissem et al., 1986b). Most of the chloroplast protein coding genes are in polycistronic transcription units (Ohyama et al., 1986; Shinozaki et al., 1986). Post-transcriptional intercistronic processing occurs for several of the polycistronic mRNAs encoding components of the photosynthetic apparatus (Berends et al., 1987; Hudson et al., 1987; Barkan, 1988; Westhoff and Herrmann, 1988), and apparently follows a conserved pathway (Gruissem et al., 1988). The possibility that the ribosomal protein mRNA processing pathways observed in *Euglena* are conserved among chloroplasts from different species awaits further studies.

The major bands detected on the RNA gel blots (Figures 3A, 5, and 6A) correspond to fully spliced mRNAs. Some of the complexity and the minor bands detected by RNA gel blot hybridization could have been due to unspliced mRNA precursors. The presence of 15 introns could result in a very complicated mixture of pre-mRNAs. Unspliced precursor mRNAs have been detected for the highly expressed *Euglena* chloroplast *psbA* gene (Hollingsworth et al., 1984) and for several other higher plant chloroplast genes (Posno et al., 1986; Westhoff and Herrmann, 1988; Barkan, 1989). Unspliced precursors have also been detected (Fukuzawa et al., 1987) for the liverwort chloroplast *petB* and *petD* loci. However, intron sequences were not detected in the cDNA sequence analysis of *Euglena* chloroplast *rps19*, *rpl16*, and *rps8* (Christopher and Hallick, 1989). This is one indication that the majority of the ribosomal protein transcripts present in the light-grown *Euglena* plastid mRNA population are fully spliced. The second line of evidence comes from the RNA gel blot hybridization experiments with intron-specific probes (D.A. Christopher and R.B. Hallick, unpublished data). Very rare mRNAs were detected, but their sizes did not correspond to the more abundant mRNAs presented here. The ratio of spliced versus unspliced transcripts seems to vary according to the developmental state of the plastid (Barkan, 1989). It has been suggested that plastid ribosomal protein mRNAs may be preferentially spliced and translated in root plastids (Deng and Gruissem, 1988; Barkan, 1989). The numerous introns present in the *Euglena* chloroplast protein genes and the ability to adapt to highly diverse growth conditions make *Euglena* particularly suitable to studies on the developmental and environmental effects on the levels of spliced versus unspliced mRNAs.

METHODS

Construction of Gene-Specific Recombinant DNAs

To create gene-specific recombinant plasmids, the exonuclease III deletion subcloning procedure (Henikoff, 1984) was utilized, with modifications for *Euglena* chloroplast DNA as previously described (Christopher and Hallick, 1989). The extent of the deletions was monitored by the use of restriction digestion and DNA sequencing (Sanger et al., 1977) of the appropriate regions. All recombinant plasmid DNAs utilized for riboprobe synthesis were verified via DNA sequence analysis to contain only the DNA inserts designated in Table 2 and Figures 3A, 5, 6A, and 7.

Chloroplast RNA Isolation

Chloroplasts were isolated in both the presence and absence of the RNase inhibitor aurintricarboxylic acid (Hallick et al., 1977) from photoautotrophically grown *Euglena gracilis* Klebs strain Z pringsheim cells as described (Hallick et al., 1982). Purified chloroplasts were resuspended in lysis buffer (0.5% SDS, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 5 mM DTT). Nucleic acid was extracted exactly according to Christopher and Hallick (1989). Remaining DNA was removed by digestion with RQ1 RNase-free DNase (Promega Biotec) at 1 unit/μg nucleic acid, 15 min, 37°C. After phenol extraction and ethanol precipitation as above, chloroplast RNA was used in RNA gel blot hybridizations and enzymatic reactions.

Probe Synthesis, RNA Gel Blot, and S1 Nuclease Analyses

RNA gel blot hybridization probes consisted of synthetic ^{32}P-labeled RNAs complementary to transcripts from the ribosomal protein genes and the open reading frames 214 and 302. The radiolabeled RNAs were prepared from linearized plasmid clones containing the gene-specific chloroplast DNA inserts listed in Table 2 and shown in Figures 3, 6, and 7. T3 and T7 RNA polymerases were used to catalyze the synthesis of radiolabeled RNAs as described in Melton et al. (1984). The sizes of the RNAs were examined by electrophoresis in 6% polyacrylamide/8 M urea sequencing gels (80 mM Tris borate, pH 7.8). In addition, the oligo-deoxynucleotide with the sequence 5′-TCGAACCAGCGACGTCCTTCCGGAGCGAGATTATGAGTCCCGTGCTATCGTCCAC-3′ was synthesized at the University of Arizona Macromolecular Structure Facility. This 55-base oligomer is complementary to, and contained entirely within, tRNA-Ile (CAU). It was ^{32}P-labeled by using T4-polynucleotide kinase (Maniatis et al., 1982). The oligomer and the synthetic radiolabeled RNAs were used as probes on filter (RNA gel) blots of chloroplast RNA (10 μg to 12 μg) fractionated on 0.66 M formaldehyde/1% agarose gels (Fourney et al., 1987). Sizes of chloroplast RNAs were determined by electrophoresing two sets of synthetic RNA size

Table 2. Gene-Specific Plasmid Clones Used for Probe Synthesis

Plasmid	Probes	*Euglena* Chloroplast Loci	Nucleotide Coordinates	Sites Used
pEZC270.11	A	*psbF*, *psbL*, *orf42*	1322–2467	HindIII T7
pEZC588.9-7	B	*rpl23*	596–861	HindIII T7
pEZC590.7	C	*rpl2*	1151–1529	EcoRI T3
pEZC591.34	D	*rps19*,3′-rpl2	1802–1980	HindIII T7
pEZC517.9-2.9	E	*rpl22*	2669–3205	EcoRI T3
pEZC517.9-3.1	F	*rps3*	3872–4584	EcoRI T3
pEZC949.9-3	*orf214*	*orf214*	NP[a]	HindIII T7
pEZC541.2.11-36	*orf302*	*orf302*	NP	EcoRI T3
pEZC541.2.4-4	K	*rpl16*	137–1078	BglII T3
pEZC942.20.3-2	L	*rpl14*	1614–1790	EcoRI T3
pEZC942.4-8.5	M	*rpl5*	1823–2367	PvuII T7
pEZC942.4-4	N	*rps8*	2514–3575	HindIII T7
pEZC948.5	O	*rpl36-trnI-rps14*	2546–3820 1–164	EcoRI T3
pEZC800.7-15	P	*rps14-trnF*	457–656	EcoRI T3
pEZC800.7-16	Q	*trnC-3′ spacer*	688–NP	BalI T3
pEZC948.42	R	*rps2*	NP	ClaI T3

[a] NP, not published.

The recombinant plasmid DNAs and their gene-specific inserts were utilized for probe synthesis in the RNA gel blot and S1 nuclease analyses (Figures 3, 5, 6, and 7). The nucleotide coordinates refer to the sequence data for probes A through F presented in Christopher et al. (1988) and for probes K through Q in Nickoloff et al. (1989) and Christopher and Hallick (1989). Also included are the restriction sites (for plasmid linearization) and promoters (T3 or T7) used in radiolabeled RNA synthesis.

standards (Bethesda Research Laboratories). The 16S (1.5 kb) and 23S (2.8 kb) chloroplast rRNAs were additional internal size markers. Separated RNAs were transferred by diffusion to GeneScreen filters (Du Pont) and cross-linked by a 5-min exposure to ultraviolet irradiation. The filters were cut lengthwise between lanes such that several individual single-lane replica blots were obtained. Each slot blot consisted of a dilution series of chloroplast RNA and one of the plasmid clones specific for *rpl16* (Table 2) or *psbA* (pEZC514.9, Hollingsworth et al., 1984). Prehybridization and hybridization were done at 55°C as previously described (Christopher and Hallick, 1989). Hybridization and membrane wash solutions using the oligomer probe were as above except the temperature was 37°C. Autoradiography was done with 12-hr to 16-hr exposures on Kodak X-Omat AR x-ray film. Control hybridizations included RNA probes complementary to each of the Bluescribe vector strands, which produced no signals to *Euglena* chloroplast RNA under the above conditions.

S1 Nuclease and Primer Extension Analyses

The S1 nuclease protection experiments utilized a uniformly radiolabeled single-stranded RNA probe of 200 nt spanning exon 2 of *rps14*, the 3′ intergenic spacer, and *trnF*. It was synthesized from plasmid pEZC800.7-15 listed in Table 2 (probe shown in Figure 7). Before S1 nuclease digestion, the probe was annealed either to chloroplast RNA or to the denatured plasmid pEZC800.7-15. The S1 nuclease digestions were performed as described in Quarless and Heinrich (1986) and analyzed on 6% polyacrylamide/8 M urea DNA sequencing gels.

The purified oligo-deoxynucleotide primer 5′-CTTAAACTGATCATAAAATT for *rpl23* exon 1 (positions 275 to 294, in Christopher et al., 1988) was synthesized and 5′-end labeled as described above. The annealed primer/RNA mixture was extended using reverse transcriptase as previously described (Christopher and Hallick, 1989). Additional control primer extension reactions were done in parallel except that the 5′-^{32}P-labeled primer was annealed to 1 μg of a genomic chloroplast clone containing the *rpl23* gene, a single-stranded recombinant phagemid, pEZC270 (Christopher et al., 1988).

ACKNOWLEDGMENTS

We wish to thank Dr. Gloria Yepiz-Plascencia for assistance with the S1 nuclease analysis experiments, and Jane Huff and Noreen Sleator. This work was supported by a grant to R.B.H. from the National Institutes of Health.

Received January 29, 1990; revised May 7, 1990.

REFERENCES

Barkan, A. (1988). Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. EMBO J. **7**, 2637–2644.

Barkan, A. (1989). Tissue-dependent plastid RNA splicing in

maize: Transcripts from four plastid genes are predominantly unspliced in leaf meristems and roots. Plant Cell **1,** 437–445.

Berends, T., Gamble, P.E., and Mullet, J.E. (1987). Characterization of the barley chloroplast transcription units containing *psaA-psaB* and *psbD-psbC*. Nucl. Acids Res. **15,** 5217–5240.

Cerretti, D.P., Dean, D., David, G.R., Bedwell, D.M., and Nomura, M. (1983). The *spc* ribosomal protein operon of *E. coli*: Sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucl. Acids Res. **11,** 2599–2616.

Christopher, D.A., and Hallick, R.B. (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. Nucl. Acids Res. **17,** 7591–7608.

Christopher, D.A., Cushman, J.C., Price, C.A., and Hallick, R.B. (1988). Organization of ribosomal protein genes *rpl23*, *rpl2*, *rps19*, *rpl22* and *rps3* on the *Euglena gracilis* chloroplast genome. Curr. Genet. **14,** 275–286.

Clayton, D.A. (1984). Transcription of the mammalian mitochondrial genome. Annu. Rev. Biochem. **53,** 573–594.

Cushman, J.C., Christopher, D.A., Little, M.C., Hallick, R.B., and Price, C.A. (1988a). Organization of the *psbE*, *psbF*, *orf38* and *orf42* gene loci on the *Euglena gracilis* chloroplast genome. Curr. Genet. **13,** 173–180.

Cushman, J.C., Hallick, R.B., and Price, C.A. (1988b). The two genes for the P700 chlorophyll a apoproteins on the *Euglena gracilis* chloroplast genome contain multiple introns. Curr. Genet. **13,** 159–172.

Deng, X.-W., and Gruissem, W. (1987). Control of plastid gene expression during development: The limited role of transcriptional regulation. Cell **49,** 379–387.

Deng, X.-W., and Gruissem, W. (1988). Constitutive transcription and regulation of gene expression in non-photosynthetic plastids of higher plants. EMBO J. **7,** 3301–3308.

Dix, D.B., and Thompson, R.C. (1989). Codon choice and gene expression: Synonymous codons differ in translational accuracy. Proc. Natl. Acad. Sci. USA **86,** 6888–6892.

Evrard, J. L., Kuntz, M., and Weil, J. H. (1990). The nucleotide sequence of five ribosomal protein genes from the cyanelles of *Cyanophora paradoxa*: Implication concerning the phylogenetic relationship between cyanelles and chloroplasts. J. Mol. Evol., **30** (1), 16–25.

Fourney, R.M., Miyakoshi, J., Day, R.S., III, and Paterson, M.C. (1987). Northern blotting: Efficient RNA staining and transfer. Focus **10,** 5–7.

Fromm, H., Edelman, M., Koller, B., Goloubinoff, P., and Galun, E. (1986). The enigma of the gene coding for ribosomal protein S12 in the chloroplasts of *Nicotiana*. Nucl. Acids Res. **14,** 883–897.

Fukuzawa, H., Yoshida, T., Kohchi, T., Okumura, T., Sawano, Y., and Ohyama, K. (1987). Splicing of group II introns in mRNAs coding for cytochrome b_6 and subunit IV in the liverwort *Marchantia polymorpha* chloroplast genome. FEBS Lett. **220,** 61–66.

Gingrich, J.C., and Hallick, R.B. (1985). The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene. J. Biol. Chem. **260,** 16156–16161.

Greenberg, B.M., and Hallick, R.B. (1986). Accurate transcription and processing of 19 *Euglena* chloroplast tRNAs in a *Euglena* soluble extract. Plant Mol. Biol. **6,** 89–100.

Gruissem, W., and Zurawski, G. (1985). Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. EMBO J. **4,** 1637–1644.

Gruissem, W., Elsner-Menzel, C., Latshaw, S., Narita, J., Schaffer, M., and Zurawski, G. (1986a). A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription. Nucl. Acids Res. **14,** 7541–7556.

Gruissem, W., Greenberg, B.M., Zurawski, G., and Hallick, R.B. (1986b). Chloroplast gene expression and promoter identification in chloroplast extracts. Methods Enzymol. **118,** 253–269.

Gruissem, W., Barkan, A., Deng, X.-W., and Stern, D. (1988). Transcriptional and post-transcriptional control of plastid mRNA levels in higher plants. Trends Genet. **4,** 258–263.

Hallick, R.B., and Bottomley, W. (1983). Proposals for the naming of chloroplast genes. Plant Mol. Biol. Rep. **1,** 38–43.

Hallick, R.B., and Buetow, D.E. (1989). Chloroplast DNA. In The Biology of *Euglena*, D.E. Buetow, ed (New York: Academic Press) pp. 351–414.

Hallick, R.B., Chelm, B.K., Gray, P.W., and Orozco, E.M. (1977). Use of aurintricarboxylic acid as an inhibitor of nucleases during nucleic acid isolation. Nucl. Acids Res. **4,** 3055–3063.

Hallick, R.B., Richards, O.C., and Gray, P.W. (1982). Isolation of intact, superhelical chloroplast DNA from *Euglena gracilis*. In Methods in Chloroplast Molecular Biology, M. Edelman, R.B. Hallick, and N.-H. Chua, eds (New York: Elsevier Biomedical), pp. 281–294.

Hanley-Bowdoin, L., and Chua, N.-H. (1987). Chloroplast promoters. Trends Biochem. **12,** 67–70.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA splicing. Gene **28,** 351–359.

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). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217,** 185–194.

Hollingsworth, M.J., Johanningmeier, U., Karabin, G.D., Stiegler, G.L., and Hallick, R.B. (1984). Detection of multiple, unspliced precursor mRNA transcripts for the M_r 32,000 thylakoid membrane protein from *Euglena gracilis* chloroplasts. Nucl. Acids Res. **12,** 2001–2017.

Hudson, G.S., Mason, J.G., Holton, T.A., Koller, B., Cox, G.B., Whitfeld, P.R., and Bottomley, W. (1987). A gene cluster in the spinach and pea chloroplast genomes encoding one CF_1 and three CF_0 subunits of the H^+-ATP synthase complex and the ribosomal protein S2. J. Mol. Biol. **196,** 283–298.

Karabin, G.D., Farley, M., and Hallick, R.B. (1984). Chloroplast gene for M_r 32000 polypeptide of photosystem II in *Euglena gracilis* is interrupted by four introns with conserved boundary sequences. Nucl. Acids Res. **12,** 5801–5812.

Kashdan, M.A., and Dudock, B.S. (1982). The gene for a spinach chloroplast isoleucine tRNA has a methionine anticodon. J. Biol. Chem. **257,** 11191–11194.

Klein, R.R., Mason, H.S., and Mullet, J.E. (1988). Light-regulated translation of chloroplast proteins. I. Transcripts of *psaA-psaB*, *psbA*, and *rbcL* are associated with polysomes in dark-grown and illuminated barley seedlings. J. Cell Biol. **106,** 289–301.

Lindahl, L., and Zengel, J.M. (1986). Ribosomal genes in *Escherichia coli*. Annu. Rev. Genet. **20,** 297–326.

Loffelhardt, W., Kraus, M., Pfanzagl, B., Goetz, M., Brandtner, M., Markmann-Mulisch, U., Subramanian, A.R., Michalowski, C., and Bohnert, H.J. (1989). Cyanelle genes for components of the translational apparatus. In Endocytobiology. IV, P. Nardon, H.E.A. Schenk, and W. Schwemmler, eds, in press.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), pp. 90–91.

Mattheakis, L.C., and Nomura, M. (1988). Feedback regulation of the *spc* operon in *Escherichia coli*: Translational coupling and mRNA processing. J. Bacteriol. **170,** 4484–4492.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. **12,** 7035–7056.

Montandon, P.E., and Stutz, E. (1983). Nucleotide sequence of a *Euglena gracilis* chloroplast genome region coding for the elongation factor EF-Tu: Evidence for a spliced mRNA. Nucl. Acids Res. **11,** 5877–5892.

Montandon, P.E., and Stutz, E. (1984). Genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*. Nucl. Acids Res. **12,** 2851–2859.

Mullet, J.E. (1988). Chloroplast development and gene expression. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39,** 475–502.

Nickoloff, J.A., Christopher, D.A., Drager, R.G., and Hallick, R.B. (1989). Nucleotide sequence of the *Euglena gracilis* chloroplast genes for isoleucine, phenylalanine and cysteine transfer RNAs and ribosomal protein S14. Nucl. Acids Res. **17,** 4882.

Nomura, M., Morgan, E.A., and Jaskunas, S.R. (1977). Genetics of bacterial ribosomes. Annu. Rev. Genet. **11,** 297–347.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuch, H., and Ozeki, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature **322,** 572–574.

Posno, M., Vliet, A.V., and Groot, G.S.P. (1986). The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to *E. coli* ribosomal protein L16 is split by a large intron near its 5′ end: Structure and expression. Nucl. Acids Res. **14,** 3181–3195.

Quarless, S.A., and Heinrich, G. (1986). The use of complementary RNA and S1 nuclease for the detection and quantitation of low abundance mRNA transcripts. BioTechniques **4,** 434–438.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463–5467.

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). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J. **5,** 2043–2049.

Sijben-Muller, G.S., Hallick, R.B., Alt, J., Westhoff, P., and Hermann, R. (1986). Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase α-subunit. Nucl. Acids Res. **14,** 1029–1044.

Stern, D.B., and Gruissem, W. (1987). Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell **51,** 1145–1157.

Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K., and 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.

Thomas, F., Massenet, O., Dorne, A.M., Briat, J.F., and Mache, R. (1988). Expression of the *rpl23*, *rpl2* and *rps19* genes in spinach chloroplasts. Nucl. Acids Res. **16,** 2461–2472.

Thompson, W.W., and Whatley, J.M. (1980). Development of non-green plastids. Annu. Rev. Plant Physiol. **31,** 375–394.

Westhoff, P., and Herrmann, R. (1988). Complex RNA maturation in chloroplasts. Eur. J. Biochem. **171,** 551–564.

Whitfeld, P.R., and Bottomley, W. (1983). Organization and structure of chloroplast genes. Annu. Rev. Plant Physiol. **34,** 279–310.

Yepiz-Plascencia, G., Radebaugh, C.A., and Hallick, R.B. (1990). The *Euglena gracilis* chloroplast *rpoB* gene. Novel gene organization and transcription of the RNA polymerase subunit operon. Nucl. Acids Res. **18,** 1869–1878.

Zhou, D.X., Quigley, F., Massenet, O., and Mache, M. (1989). Cotranscription of the S10- and *spc*-like operons in spinach chloroplasts and identification of three of their gene products. Mol. Gen. Genet. **216,** 439–445.