The EMBO Journal vol.4 no.8 pp.1911-1919, 1985

Major light-harvesting polypeptides encoded in polycistronic transcripts in a eukaryotic alga

Peggy G.Lemaux and Arthur R.Grossman

Department of Plant Biology, Carnegie Institution of Washington, Stanford, CA 94305, USA

Communicated by J.D.Rochaix

By sequence analysis of previously identified fragments and low stringency hybridization of an identified gene for a phycobiliprotein subunit to total plastid DNA, we have identified four phycobiliprotein subunit genes in a eukaryotic alga, *Cyanophora paradoxa*. The four phycobiliprotein subunits, α and β of phycocyanin (PC) and allophycocyanin (APC), comprise the bulk of the light-harvesting complex in this alga. The α and β subunit genes encoding each phycobiliprotein (the products of which are required in a 1:1 ratio in the light-harvesting complex) are contiguous; however, the genes for different phycobiliproteins, PC and APC, are located in different regions of the genome. The two PC subunit genes are in the small single copy region of the plastid genome whereas the APC subunit genes are in the large single copy region and the two sets of phycobiliprotein genes are transcribed from opposite strands. The α and β subunits of both PC and APC are encoded in dicistronic transcripts and this arrangement may provide a mechanism by which the two subunits can be synthesized in equimolar amounts. Levels of the PC transcript are approximately five times that of the APC transcript which may reflect the relative abundance of their gene products in the phycobilisome. The 5′ ends of the transcripts for PC and APC were mapped and the regulatory regions identified. Several features of the promoter regions for these highly transcribed genes are described.

Key words: phycobiliproteins/plastid/*Cyanophora paradoxa*/light-harvesting complex/plastid promoters

Introduction

The major light-harvesting antennae in certain eukaryotic algae and prokaryotic cyanobacteria are macromolecular complexes, termed phycobilisomes (Gantt, 1981; Glazer, 1982), which although functionally analogous to the light-harvesting chlorophyll-protein complexes in higher plants, differ in structure and location (Arntzen, 1978; Hiller and Goodchild, 1981). These water-soluble structures, attached to but not embedded in the photosynthetic membranes, funnel light energy to the photosynthetic reaction centers (Wang and Meyers, 1976; Wang *et al.*, 1977). They can provide 30-50% of the light-harvesting capacity to the cell and can comprise 60% of total soluble protein (Bogorad, 1975). These levels can change dramatically in some eukaryotic algae and cyanobacteria in response to changing light conditions and this can be manifested as a change in total phycobilisome content (Ramus *et al.*, 1976) or in levels of specific components [modulated to maximize absorption of the prevailing wavelength (Bogorad, 1975)].

The phycobilisome contains both non-chromophoric (~15%) and chromophoric (~85%) polypeptides. The non-chromophoric polypeptides, termed linker proteins, hold the antennae in an ordered array (Lundell *et al.*, 1981). A high mol. wt. chromophoric polypeptide, the anchor protein, attaches the phycobilisome to the photosynthetic membranes (Redlinger and Gantt, 1982) and may function in transfer of energy from the phycobilisome to the photosynthetic reaction centers (Glazer *et al.*, 1983). Based on their spectral characteristics, the major chromophoric proteins, or phycobiliproteins, have been classified into three groups; allophycocyanin (APC), phycocyanin (PC) and phycoerythrin (PE). Each phycobiliprotein is composed of an α and β subunit, with each subunit being covalently bound to one or more tetrapyrrole chromophores (Gantt, 1981). Comparisons of the amino acid sequences among subunits in a given organism (DeLange *et al.*, 1981; Frank *et al.*, 1978; Glazer, 1977; Offner *et al.*, 1981; Offner and Troxler, 1983; Troxler *et al.*, 1975,1981; Zuber, 1983) have led to speculation that the genes encoding phycobiliprotein subunits may have evolved from duplications of a single progenitor gene (Glazer, 1977). Comparisons of the amino acid sequence of a given subunit among organisms show that the phycobiliproteins have been highly conserved through evolution. Among eukaryotic and prokaryotic species ~85-95% of the sequence is conserved (Bryant *et al.*, 1978; Freidenreich *et al.*, 1978; Offner *et al.*, 1981; Offner and Troxler, 1983; Troxler *et al.*, 1981; Williams and Glazer, 1978; Zuber, 1983)

The phycobiliprotein subunits and the anchor polypeptide are encoded in the chloroplast genome in eukaryotic algae; the linkers are nuclear-encoded (Belford *et al.*, 1983; Egelhoff and Grossman, 1983; Grossman *et al.*, 1983). Utilizing an immunological approach to screen a genomic library containing fragments from the plastid (cyanelle, chloroplast) genome of a eukaryotic alga, *Cyanophora paradoxa*, we identified the gene encoding the β subunit of PC (designated *ppc*B) (Lemaux and Grossman, 1984). Further sequence analysis of these immunoreactive clones revealed a second phycobiliprotein subunit gene for α-PC, *ppc*A. Low stringency hybridizations utilizing these two genes permitted us to identify and clone the genes for two other phycobiliprotein subunits, α- and β-APC (*pap*A and *pap*B, respectively). Characterization of the transcription of these genes has permitted speculation concerning the influence of their genomic organization on regulation.

Results

Identification of the genes for the α subunit of PC and the α and β subunits of APC from Cyanophora paradoxa

The gene for β-PC has been cloned and characterized recently (Lemaux and Grossman, 1984). An immunological method, utilizing antibodies to mixtures of phycobiliproteins, was used to screen *Escherichia coli* colonies containing fragments of plastid DNA from a eukaryotic alga, *C. paradoxa*. One of the immunologically reactive clones, pCPC2368, contained the nucleotide sequence for only 33 amino acids from the carboxyl-terminus of β-PC, and these were not in the same translational reading

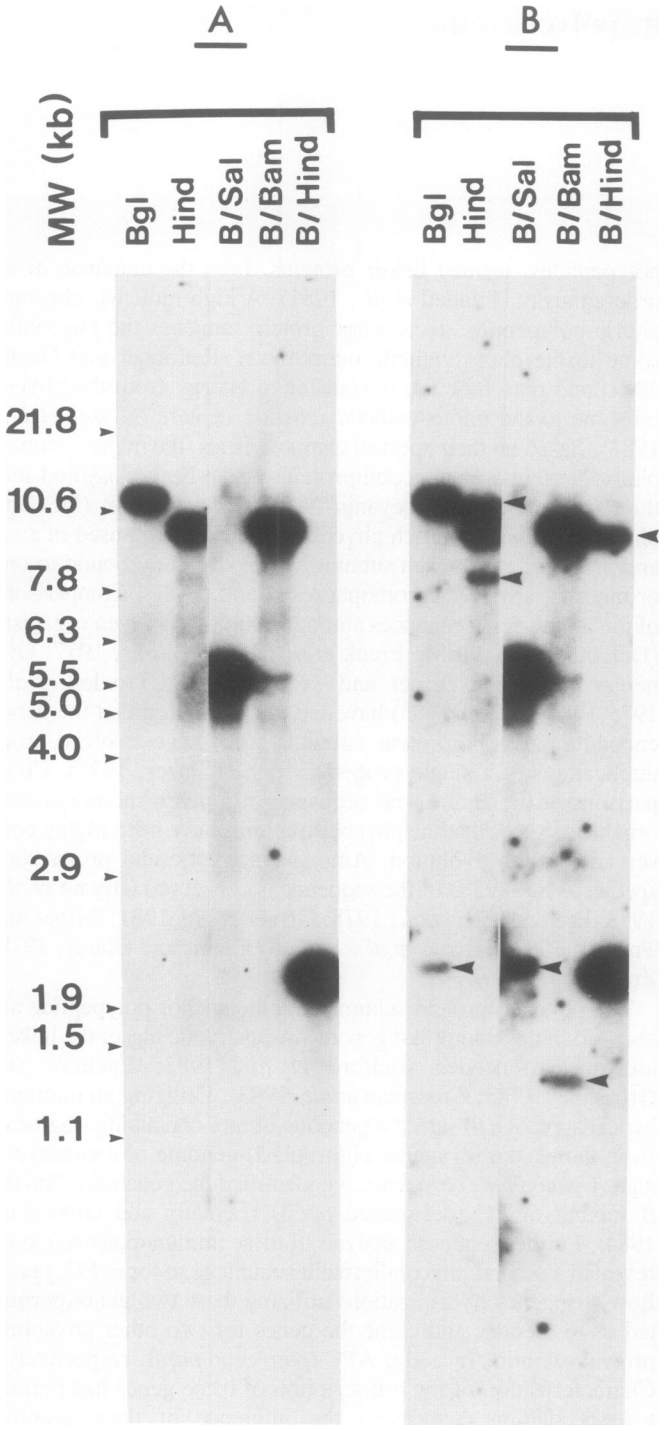

Fig. 1. Low stringency hybridizations of PC genes to total plastid DNA from *C. paradoxa*. Plastid DNA from *C. paradoxa* (~1 μg) was digested with various restriction endonucleases, electrophoresed in a 0.8% agarose-TEA gel and transferred to nitrocellulose. **(A)** Plastid DNA hybridized to a nick-translated fragment containing the entire α-PC gene and the sequence for only the carboxyl-terminal 31 amino acids of β-PC (2 x 10^6 Cerenkov c.p.m./filter). **(B)** Plastid DNA hybridized to a nick-translated fragment encoding most of β-PC (except the amino-terminal 17 amino acids) and the amino-terminal one-third of α-PC (2 x 10^6 Cerenkov c.p.m./filter). Hybridizations were at 50°C overnight. Washing in **A** was 3 x 10 min (room temperature) in 0.01 M phosphate buffer; washing in **B** was 3 x 10 min (room temperature) in 0.01 M phosphate buffer and 1 x 10 min in 0.05 M PO_4 buffer at 42°C. Exposures were for 18 h. Mol. wt. standards (left side) were pRR12 digested with *Eco*RI (Tanaka *et al.*, 1976). Arrows in **B** demarcate weakly hybridizing bands not seen in **A**. The abbreviations are Bgl, *Bgl*II; Hind, *Hind*III; B/Sal, *Bgl*II/*Sal*I; B/Bam, *Bgl*II/*Bam*HI; and B/Hind, *Bgl*II/*Hind*III.

frame as the β-galactosidase gene into which it was inserted. Therefore, we speculated that the immunoreactivity of the clone was due to expression of another phycobiliprotein subunit gene contained on that insert. Further sequencing of pCPC2368 revealed the α-PC gene downstream from the β-PC gene.

Amino acid analyses of phycobiliprotein subunits from other species have demonstrated limited homology among the various subunits (~30%), suggesting that all of these genes evolved from a single progenitor gene (Glazer, 1977). Although overall amino acid homologies are low, more detailed analyses show limited stretches which share as much as 75% homology. Therefore, to locate other phycobiliprotein subunit genes we hybridized probes encoding either mostly α-PC or mostly β-PC to digests of the *C. paradoxa* plastid genome under conditions of low stringency (allowing 45% mismatch). Both probes hybridized to the homologous fragments in the plastid genome, giving strong hybridization signals (Figure 1A and B). In addition, however, the β-PC probe hybridized weakly to other bands (Figure 1B). One of these bands, a 1.4-kb *Bgl*II/*Bam*HI fragment, was cloned into the *Bam*HI site of pUC8 yielding the plasmid pCPC1422. Restriction maps for both this fragment and the PC gene cluster were generated (Figure 2A and B).

Portions of the 1.4-kb insert of pCPC 1422 were sequenced and found to contain the genes for the α and β subunits of APC. Nucleotide and predicted amino acid sequences for the amino-terminal 15 amino acids of the four phycobiliprotein subunits identified in *C. paradoxa* are also shown in Figure 2. For comparison the amino-terminal sequences of homologous subunits (determined by sequencing of polypeptides) are shown for another eukaryotic alga, *Cyanidium caldarium* (Offner and Troxler, 1983). A strong conservation of amino acid sequence between phycobiliprotein subunits of these two species is seen and this was expected since the same subunit isolated from eukaryotic and prokaryotic species exhibits 85–95% amino acid conservation (Bryant *et al.*, 1978; Freidenreich *et al.*, 1978; Offner *et al.*, 1981; Offner and Troxler, 1983; Troxler *et al.*, 1981; Williams and Glazer, 1978; Zuber, 1983).

A search in the regions upstream from the start of the structural genes reveals sequences complementary to the 3′ end of the 16S rRNA from *E. coli* (Shine and Dalgarno, 1974) which are 9–11 bases upstream from the sites of translation initiation. The spacing between the β- and α-PC genes is 81 bp, while that for the two APC subunit genes is 42 bp. Stop codons are present in all frames between both genes. The absence of features characteristic of transcriptional initiation sites in these intergenic regions (Hawley and McClure, 1983; Rosenberg and Court, 1979) is consistent with the analysis of transcriptional data presented below.

Mapping of the APC genes on the plastid genome of C. paradoxa

A restriction map of the plastid genome is shown in Figure 3. The position of the β-PC gene in the small single copy region and the direction of its transcription have been previously reported (Lemaux and Grossman, 1984). Based on restriction mapping and nucleotide sequence data of previously identified immunologically reactive clones the α-PC gene, *ppc*A, can now be placed on the map near the β-PC gene, *ppc*B. The map position of the APC genes was determined by hybridization of the 1.4-kb insert from pCPC1422, containing both APC subunits, to various restriction digests of *C. paradoxa* plastid DNA. Species hybridizing were a >30-kb *Bam*HI fragment, a 2.1-kb *Bgl*II fragment, an ~30-kb *Sal*I fragment, a >50-kb *Sma*I fragment and an 8.8-kb *Xho*I fragment (data not shown). Based on published restriction maps of the *C. paradoxa* plastid genome (Bohnert *et*

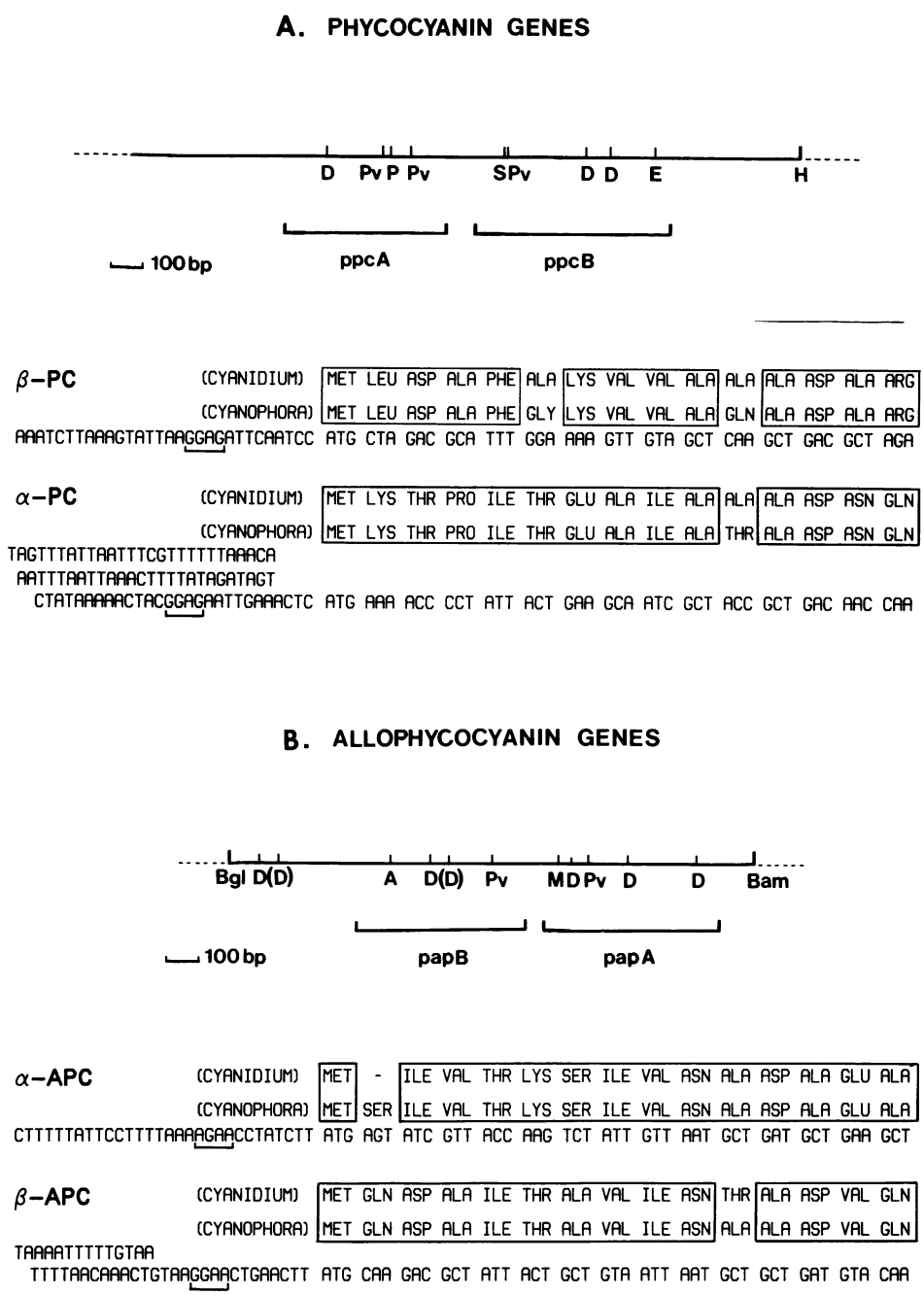

Fig. 2. Restriction enzyme maps of the PC and APC genes from *C. paradoxa* and nucleotide sequences for their 5′ ends and predicted amino acid sequences. **(A)** The restriction map for the PC genes was determined by analysis of two fragments subcloned into pUC8 [the *Eco*RI-*Hind*III fragment (0.45 kb) and the *Eco*RI-*Pst*I fragment (0.8 kb) from one of the plasmids identified by immunological screening pCPC4067 (Lemaux and Grossman, 1984)] and of pCPC2368 (Lemaux and Grossman, 1984) which contains the sequences from the *Sau*3A site leftward. Subsequent nucleotide sequence analysis permitted assignment of individual genes to certain fragments. *ppc*B is the designation for the gene encoding β-PC; *ppc*A, the designation for the gene encoding α-PC. The nucleotide sequence encoding the amino terminus of β-PC and the pregene region was determined by labeling the *Eco*RI site with polynucleotide kinase and by fill-in with Klenow enzyme. The amino acid sequence predicted from the nucleotide sequence is shown after (CYANOPHORA). For comparison the amino acid sequence determined from purified phycobiliproteins of *C. caldarium* (Offner and Troxler, 1983) is shown after (CYANIDIUM). Amino acids in common between the two squences are boxed. A possible ribosome binding site is underlined (—). The nucleotide sequence of the region encoding the amino terminus of α-PC and of its intergenic region was obtained using clone pCPC2368 (Lemaux and Grossman, 1984) by labeling (with polynucleotide kinase and Klenow enzyme) at the *Eco*RI site of the vector (close to the *Sau*3A site shown in the restriction map). The entire intergenic region between the β-PC and α-PC genes is shown; 'TAG' being the terminator codon for β-PC. **(B)** The restriction map for the APC genes was determined by analysis of a 1.4-kb *Bgl*II/*Bam*HI fragment cloned from the plastid genome into pUC8 (pCPC1422). The location of the *Dde*I site in parentheses (D) is not known but is in one of the two locations indicated. Subsequent nucleotide sequence analysis permitted assignment of the genes to certain fragments. *pap*A is the designation for the gene encoding the α subunit of APC; *pap*B the designation for the β subunit. The nucleotide sequence of the region encoding the amino terminus of α-APC and of its pregene region was obtained by labeling the *Bam*HI-*Dde*I fragment with polynucleotide kinase, separating the strands and sequencing both. The nucleotide sequence of the region encoding the amino terminus of β-APC and of the intergenic region was derived by digestion of the 1.4-kb APC insert from pCPC1422 with *Msp*I, labeling with polynucleotide kinase and redigestion with *Ava*II. The entire intergenic region between α-APC and β-APC is shown; 'TAA' being the terminator codon for α-APC. Abbreviations: A, *Ava*I; Bam, *Bam*HI; Bgl, *Bgl*II; D, *Dde*I; E, *Eco*RI; M, *Msp*I; P, *Pst*I; Pv, *Pvu*II and S, *Sau*3a. All sites for *Ava*I and *Sau*3a are not indicated on these maps; only those used for sequencing analysis. *Msp*I sites are not shown for the PC gene cluster.

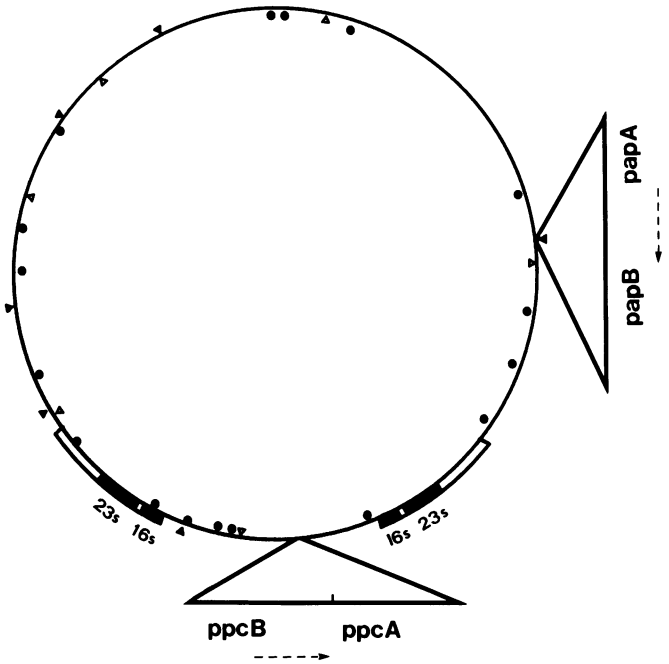

Fig. 3. Location and direction of transcription of the phycobiliprotein genes on the plastid genome of *C. paradoxa*. The plastid map has been redrawn from Kuntz *et al.* (1984) and Bohnert *et al.* (1983). The position of the β-PC gene (designated *ppc*B) has been reported previously (Lemaux and Grossman, 1984). The position of the α-PC gene (designated *ppc*A) and the direction of transcription (--->) was inferred from the position assigned to the β-PC gene, after sequence analysis showed the relationship of these two genes. The map position of the genes for the α and β subunits of APC (designated *pap*A and *pap*B, respectively) was determined by Southern hybridization of APC gene fragments to plastid DNA cut with *Bam*HI, *Xho*I, *Sal*I, *Bam*HI/*Sal*I and *Bam*HI/*Xho*I. Sizes of hybridizing bands were then correlated with published fragment sizes. The direction of transcription (--->) was inferred from the sequence of the genes, the presence of a *Bam*HI site on one end of the gene cluster, and the sizes of hybridizing framgents in *Bam*HI/*Sal*I and *Bam*HI/*Xho*I digests. The bracketed areas at the bottom of the map represent the inverted repeat regions containing the 16S and 23S rRNA cistrons. Symbols are (△) inner circle, *Sal*I, (▲) outer circle, *Bam*HI, (●), *Xho*I.

al., 1983; Kuntz *et al.*, 1984), the only map position consistent with these data is in the large single copy region (Figure 3). Since the orientation of the α and β subunit genes (*pap*A and *pap*B) on the 1.4-kb *Bam*HI/*Bgl*II fragment was known (Figure 2B), the direction of transcription of the genes could be determined. Hybridization of the 1.4-kb fragment to a 1.6-kb *Bam*HI/*Sal*I plastid DNA fragment and a 6.0-kb *Bam*HI/*Xho*I fragment (Figure 3) permitted determination of the orientation of the genes on the plastid genome. The direction of transcription of the APC genes is opposite to that of the PC genes and therefore they are transcribed from the opposite strand.

RNA analysis of PC and APC genes

Northern analyses were performed using radioactive phycobiliprotein genes to hybridize to poly(A)$^-$ RNA from *C. paradoxa* (Figure 4). An α-PC-specific probe encoding the carboxyl-terminal two-thirds of α-PC hybridized to a transcript of ~1350 bases (Figure 4A). The same result was obtained using a β-PC-specific probe containing the nucleotide sequence covering the 5′-untranslated region and encoding the amino-terminal 17 amino acids of β-PC (data not shown). No hybridization between different phycobiliprotein subunit genes occurred under the conditions of stringency used in the Northern analyses (25%

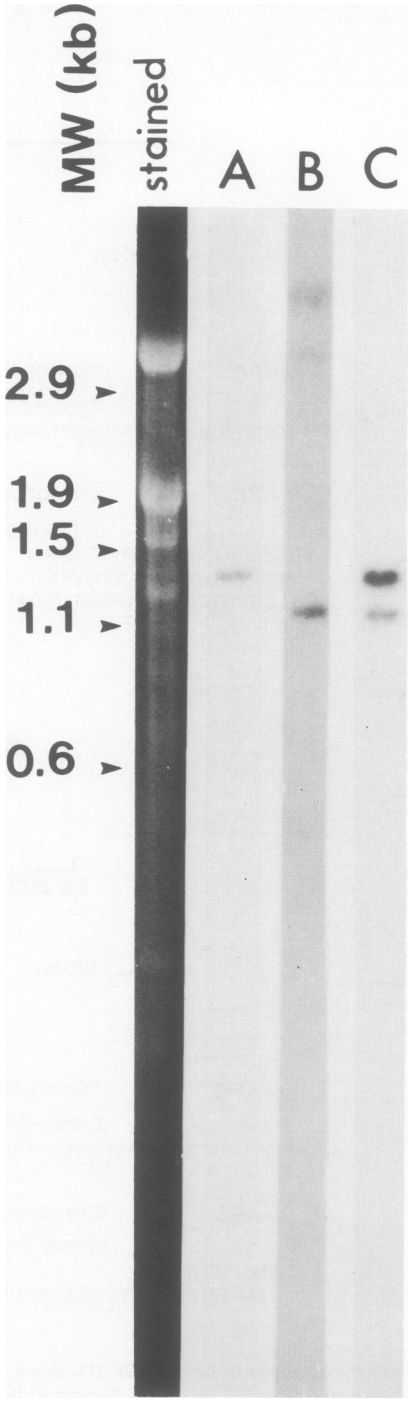

Fig. 4. Northern analyses of phycobiliprotein transcripts from poly(A)$^-$ RNA of *C. paradoxa*. Poly(A)$^-$ RNA was isolated from *C. paradoxa* as described and 3 μg was loaded per lane. One lane was used for staining; identical lanes were used for the hybridizations. **(A)** Hybridization to a fragment encoding the carboxyl-terminal two-thirds of α-PC (leftward from *Pst*I site in Figure 2A). **(B)** Hybridization to the entire *Bam*HI-*Bgl*II fragment (Figure 2B) containing both the α-APC and β-APC genes. **(C)** Hybridization simultaneously to the PC and APC probes described in **A** and **B**. For all hybridizations 500 000 Cerenkov c.p.m. of each probe per filter was used; hybridizations were at 65°C; and washing was 3 x 10 min at room temperature in 0.01 M phosphate buffer. Exposures were for 10 h. Mol. wt. standards, a mixture of pRR12 digested with *Eco*RI (Tanaka *et al.*, 1976) and pBR322 digested with *Hpa*II, were treated identically to the RNA samples and stained for visualization.

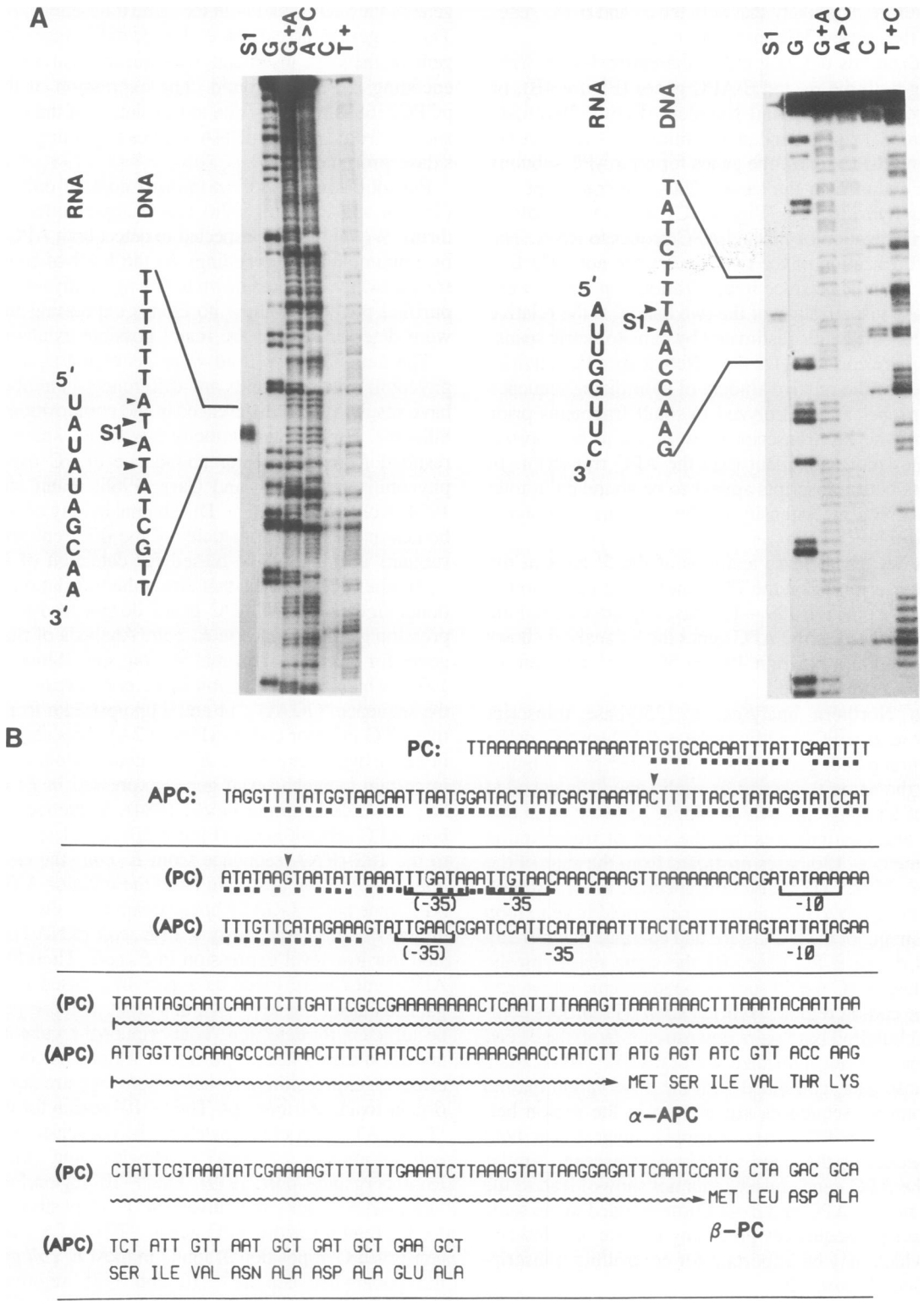

Fig. 5. S1 nuclease and nucleotide sequence analyses of the 5′-regulatory regions of the PC and APC genes. **(A) Left.** Mapping of the 5′ end of the PC transcript by S1 nuclease digestion. On the right of this portion of the figure is the sequencing ladder of the *Eco*RI-*Hind*III fragment encoding the amino terminus of β-PC (see Figure 2A); the five different chemistries performed are as indicated. The lane marked S1 contains the S1 nuclease reaction performed as described in Materials and methods. The nucleotide sequence of the region surounding the S1 nuclease-resistant fragments is given to the left of the reaction (DNA), with arrows indicating the ends of the protected fragments; the size of the arrow is indicative of the amount of that size fragment seen. The sequence of the 5′ end of the transcript is also shown (RNA). **Right.** Mapping of the 5′ end of the APC transcript by S1 nuclease digestion. On the right of this portion of the figure is the sequencing ladder of the *Dde*I-*Bam*HI fragment encoding the amino terminus of α-APC (see Figure 2B). The format for the remainder of this portion of the figure is as described above. **(B)** Nucleotide sequences were determined as described in the legends to Figures 2A and 2B, with the PC sequence being uppermost. The sequences are aligned so that the start points of transcription for the PC and APC transcripts are at the same position indicated by (↝). The arrows continue through the 5′-untranslated regions to the beginning of the coding sequences (indicated by α-APC or β-PC). Possible '−10' and '−35' regions are bracketed and labeled as '−10, '−35' and '(−35)'. A large region of dyad symmery in the non-transcribed portion of each gene cluster is indicated with symbols (■) under complementary bases. Its axis of symmetry is marked (▼).

mismatch). Therefore, it is likely that both the β- and α-PC genes are encoded in the same 1350-base transcript.

Similar hybridizations utilizing either the entire 1.4-kb fragment, containing both the α- and β-APC genes (Figure 4B), or an α-specific probe [*Pvu*II-*Bam*HI fragment (Figure 2B)] (data not shown) revealed a single band at a position of ~1200 bases. As with the α- and β-PC genes, the genes for both APC subunits are likely to be encoded in the same 1200-base transcript.

The remaining hybridization (Figure 4C) shows the result of simultaneous hybridization of PC and APC probes to RNA. The intensities of hybridization of the two species are not reflective of their relative abundance since hybridization conditions were optimized for easy visualization of the two bands. If the relative intensities of these two bands [estimated by densitometric scans, (data not shown)] are corrected for the different specific activities of the probes and for the relative amount of hybridizing sequence in each probe (probes are hydrolyzed to small fragments prior to hybridization), the PC transcript is estimated to be approximately five times more abundant than the APC transcript. In absolute amounts both transcripts appear to be abundant in total poly(A)$^-$ RNA given the intensity of the hybridization signal from a 10 h exposure.

Figure 5A shows S1 nuclease analyses of the 5′ ends of the PC and APC transcripts. For the PC genes these data show a cluster of bands at a position 129 – 133 bases upstream from the ATG of the β-PC gene. For the APC genes the S1 analysis shows two bands protected at a position 49 and 50 bases upstream of the ATG for the α-APC gene.

According to Northern analyses, a 1350-base transcript hybridizes either to an α-PC-specific probe, a β-PC-specific probe or a probe containing both genes. Therefore, either both subunits are encoded in the same transcript or each subunit is encoded in a transcript of a similar size which cannot be resolved in this gel system. S1 analysis indicates that the start of transcription for the β-PC gene is ~130 bases upstream from the start of the coding sequence. If the β-PC subunit is encoded in a 1350-base transcript different from α-PC this transcript would have to loop out near the beginning of the α-PC gene and continue downstream past the 3′ end of the α-PC gene. By the same reasoning the transcript encoding α-PC would start in the intergenic region and continue past the coding region. Hybridization with probes commencing several hundred base pairs downstream from the 3′ end of the α-PC gene did not hybridize with poly(A)$^-$ RNA (data not shown). These data, combined with the observation that no *E. coli*-like promoter sequences are present in the region between the β-PC and α-PC genes, strongly suggest that both subunits are encoded in the same 1350-base transcript. Similar analyses with the APC gene cluster (data not shown) lead to the conclusion that both α-APC and β-APC are encoded in the same 1200-base transcript. Sequences preceding the site of transcription initiation which may be important for controlling transcription are indicated (Figure 5B).

Discussion

Expression of phycobiliprotein subunits in E. coli

The original immunological screen which permitted us to identify and characterize the β-PC gene depended on the expression of plastid genes in *E. coli* (Lemaux and Grossman, 1984). Since immunoreactive clones were found with inserts in both orientations in the plasmid, we concluded that β-PC was probably being expressed from its own promoter. In one plasmid, pCPC2368, however, only the sequence encoding the carboxyl-terminal 33 amino acids of β-PC was found (contiguous to the β-galactosidase gene of the vector but not in the same translational reading frame). This suggested the existence of a second phycobiliprotein subunit gene on the same insert and, by sequence analysis, a second gene encoding α-PC was found. The expression of this gene from pCPC2368 is probably due to translation of the phycobiliprotein subunit from a fused mRNA species beginning at the β-galactosidase promoter.

Phycobilisomes of *C. paradoxa* contain only PC and APC (Trench and Ronzio, 1978) (some algae contain no phycoerythrin). We might have expected to detect both APC and PC genes by immunological screening. At the level of sensitivity of our screen (which detected down to 0.1 ng of phycobiliproteins from purified phycobilisomes) no clones expressing the APC genes were detected. There are many possible exlanations for this.

The antibodies we used were generated against mixtures of phycobiliprotein subunits and differences in antibody titers may have resulted from differences in the composition of the phycobilisome. Higher titer antibody toward PC subunits could have resulted from the greater abundance of PC over APC in the phycobilisome (Craig and Carr, 1968; Gantt and Lipschultz, 1974; Köller *et al.*, 1977). Differential avidity of the various antibodies or varying antigenicity of the different phycobiliprotein subunits also may have biased for detection of PC genes.

Another possibility is that either the translational or transcriptional signals for the APC genes do not lead to detectable expression of these genes in *E. coli*. Analysis of the α- and β-PC genes for possible ribosomal binding sites (Shine and Dalgarno, 1974) which might function in *E. coli* revealed that both have the sequence 'GGAG', 11 and 9 bp upstream from their respective ATG initiator codons (Figure 2A). This sequence is present in many highly expressed *E. coli* genes (Gold *et al.*, 1981) and in certain cyanobacterial genes expressed in *E. coli* (Fisher *et al.*, 1981; de Lorimier *et al.*, 1984). Sequences upstream from both APC subunit genes (Figure 2B) have less complementarity to the 16S rRNA sequence from *E. coli*; the α-APC gene has only a 'GA' 9 bp upstream from the initiator ATG while the β-APC gene has a 'GGA' 9 bp upstream from the ATG. This lower degree of complementarity with *E. coli* rRNA could be responsible for low level expression in *E. coli*. Therefore, even if the APC genes were made as a part of a fused message with β-galactosidase, the level of translation of these genes might not be sufficient for detection. Transcriptional regulatory regions contained on the cloned *C. paradoxa* fragments could also have an effect on expression in *E. coli* where they are not part of a fused β-galactosidase transcript. The '−10' region for the APC genes, 'TATTAT', is a close match to the consensus sequence for *E. coli* promoters, 'TAtaaT' (Hawley and McClure, 1983; Rosenberg and Court, 1979). The '−10' region for the PC genes, 'TATAAA', lacks the 'invariant T' at position 6. A possible '−35' region for the APC genes, 'TTCATA', lacks the highly conserved G at position 3, found in most *E. coli* genes; however, the spacing between the '−10' and '−35' regions, 18 bp, is optimal for expression in *E. coli*. A possible '−35' region for the PC genes, 'TTGATA', is a close match to the consensus *E. coli* sequence 'TTGaca', but the distance between the '−10' and '−35' regions, 23 bp, is longer than that seen for any *E. coli* gene characterized (Hawley and McClure, 1983).

Identification and genomic organization of phycobiliprotein subunit genes

Since phycobiliprotein subunits are a family of polypetides which share amino acid homology (up to 75% for limited regions) (DeLange and Williams, 1981; Frank *et al.*, 1978; Glazer, 1977; Offner *et al.*, 1981; Offner and Troxler, 1983; Troxler *et al.*,

1975,1981; Zuber, 1983), a gene for one subunit might hybridize at low stringencies to other phycobiliprotein subunit genes. Under conditions allowing 45% mismatch, a predominantly β-PC probe hybridized to the APC genes whereas a predominantly α-PC probe did not. This difference might be related to the fact that [in another eukaryotic alga, *C. caldarium* (Offner and Troxler, 1983)] β-PC has greater overall amino acid homology with α-APC (36%) and β-APC (36%) than does α-PC (31 and 26%, respectively). More detailed comparisons in *C. caldarium* reveal that β-PC has an extensive region of amino acid homology with β-APC (19 of 25 amino acids) which is not seen in other comparisons. Therefore, the weak hybridization signal seen with the β-PC probe probably results from hybridization of this gene to the β-APC gene.

The genes for the α and β subunits of each individual phycobiliprotein are clustered together on the plastid genome, two in the large single copy region (*pap*A and *pap*B) and two in the small single copy region (*ppc*A and *ppc*B). This is the first assignment of a structural gene to the small single copy region of any plastid genome (Schreier and Bohnert, 1985; Whitfeld and Bottomley, 1983). The transcription of these two sets of genes is in opposite directions, so the transcripts are synthesized from different strands of the plastid genome. In the case of PC, the β subunit gene is found at the 5′ end of the α subunit gene, while in the case of APC, the α-APC gene is at the 5′ end of the β-APC gene. This arrangement appears to have been conserved through evolution as a β-PC gene has been found at the 5′ end of an α-PC gene in two prokaryotic cyanobacterial species, *Agmenellum quadruplicatum* (de Lorimier *et al.*, 1984; Pilot and Fox, 1984) and *Fremyella diplosiphon* (Conley *et al.*, 1985).

Transcriptional analysis in C. paradoxa

Northern analyses coupled with S1 nuclease analysis indicate that the α and β subunits of PC are encoded in the same transcript, of ~1350 bases. By extrapolation from amino acid sequence data from other organisms (Frank *et al.*, 1978; Offner *et al.*, 1981; Troxler *et al.*, 1975,1981), β-PC has ~170 amino acids, requiring 510 bp of coding sequence. The α-PC polypeptide is ~160 amino acids, requiring 480 bp of coding sequence. Nucleotide sequence analysis reveals an intergenic region of 81 bp while S1 analysis reveals a 5′-untranslated region of ~130 bp. This leaves ~150 bp in the 3′-untranslated region of the PC transcript. The APC transcript was sized at 1200 bases and this smaller size is reflected in a shorter 5′-untranslated region (50 bases) and a shorter intergenic region (42 bases). Based on determinations in other prokaryotic and eukaryotic organisms (DeLange *et al.*, 1981; Offner and Troxler, 1983; Zuber, 1983), the α and β subunits of APC are ~160 amino acids each, requiring a total of 960 bp of coding sequence. As with the PC transcript, this leaves ~150 bases for the 3′-untranslated region.

The relative amounts of PC and APC in the phycobilisome of a prokaryotic cyanobacterium have been estimated at 5:1 (Glazer *et al.*, 1983). In contrast the ratio of the α to the β subunit of any given phycobiliprotein is 1:1. Therefore, subunits of the same phycobiliprotein need to be present in equimolar amounts while different phycobiliproteins are needed in very different amounts. The arrangement of the phycobiliprotein subunit genes on this plastid genome provide a possible mechanism by which the cell might accomplish this type of regulation. The genes for the subunits of a single phycobiliprotein are tandemly arranged on a transcript and, therefore, could be synthesized in approximately equimolar amounts. Such a translational coupling mechanism has been proposed to ensure equimolar synthesis of the β and ε subunits of ATPase in spinach, which are also encoded in a dicistronic mRNA (Zurawski *et al.*, 1982). Both subunits for different phycobiliproteins are on separate transcripts and hence can be more readily differentially regulated.

Levels of the individual transcripts for the PC and APC genes are consistent with the differential levels of their gene products present in the phycobilisomes of other species (Glazer *et al.*, 1983; Siegelman and Kycia, 1982). (The ratio of APC:PC in *C. paradoxa* has not been accurately measured). The degree of hybridization of the two mRNA species to gene-specific probes (Figure 4C), when corrected for specific activities and the proportion of gene-specific sequences in the probe, indicates a ~5-fold greater abundance of the PC transcript over the APC transcript.

Differences in levels of PC and APC transcripts could be the result of differential rates of degradation or synthesis of the mRNA. In another photosynthetic bacterium the differential stability of specific regions of a polycistronic transcript accounts for the differences in levels of certain light-harvesting and reaction center polypeptides (Belasco *et al.*, 1985). Nothing, however, is known about message stability in *C. paradoxa*. Potential hairpin structures vary markedly between the two transcripts. The 5′-non-translated and intergenic regions of the PC transcript have several almost contiguous hairpin structures (length 6–20 bases; loopouts 3–13 bases) while the same regions of the APC transcript have few. If this secondary structure is relevant to differential transcript stability (Movva *et al.*, 1980), the longer 5′-untranslated and intergenic regions of the PC transcript may be a necessarily conserved feature.

Features of transcriptional and translational regulatory sequences

Prokaryotic cyanobacteria are believed to be the progenitors of the eukaryotic chloroplast (Gray and Doolittle, 1982; Margulis, 1970). The sequence of 15 bases at the 3′ end of the 16S rRNA from one cyanobacterial species differs by only 1 base from that of *E. coli* (Borbély and Simoncsits, 1981). The sequence at the 3′ end of several chloroplast 16S rRNAs has been determined and does not differ in the terminal 8 bases from the *E. coli* sequence (Schwarz and Kossel, 1979; Tohdoh and Sugiura, 1982). Hence it is likely that the 3′ end of 16S rRNA from *C. paradoxa* will be similar to that of *E. coli* and that the considerations noted previously for *E. coli* expression may apply in *C. paradoxa*.

An analysis of the 5′-regulatory regions involved in transcriptional control of the two gene sets reveals several interesting features which may be related to differential rates of synthesis in *C. paradoxa*. Key characteristics of chloroplast structural gene promoters have recently been reviewed (Schreier and Bohnert, 1984; Whitfeld and Bottomley, 1983) and most have features in common with those found in *E. coli*. The '−10' regions for the PC and APC genes are very similar, in sequence and spacing from the ATG, to characterized chloroplast '−10' regions. The only striking exception is the absence at position 6 in the PC '−10' region of the 'invariant T' found in all chloroplast genes surveyed (Schreier and Bohnert, 1984; Whitfeld and Bottomley, 1983). The '−35' regions for both genes are similar to those found in chloroplast structural genes; however the APC region does lack the highly conserved 'G' at position 3. The spacing between these two regions was different for the two gene sets but the numbers of chloroplast genes characterized to date is not sufficient to estimate optimal spacing for chloroplast genes. Because it is not known what constitutes optimal spacing in *C. paradoxa*, it is of potential importance that 31 bp upstream from the '−10' region in both APC and PC are the sequences 'TTGAAC' and 'TTGATA', respectively. Although separated from the '−10' region by a considerable distance, these sequences are

a close match to consensus chloroplast '−35' sequences. Isolation and characterization of additional genes from this organism should provide insight into the features relevant to efficient transcription.

There may be other features relevant to the transcription of the phycobiliprotein subunit genes. In one cyanobacterium, *A. quadruplicatum*, an open reading frame of 38 amino acids was found in the 5′-transcribed region upstream from the structural gene for β-PC (de Lorimier *et al.*, 1984; Pilot and Fox, 1984). This open reading frame was preceded by a potential ribosomal binding site and an ATG initiator codon and was speculated to be involved in regulation of the PC genes (Pilot and Fox, 1984). An inspection of the 130-base transcribed region at the 5′ end of the β-PC gene in *C. paradoxa* does not reveal any open frames as long as that in *A. quadruplicatum* and no open reading frame was found with characteristics of a translated polypeptide (an initiator codon and/or ribosomal binding site).

Nucleotide sequences of the regions upstream of the transcription start sites were analyzed for possible structural features which might be involved in the expression of these genes. The most striking feature is the presence of a large region of imperfect dyad symmetry preceding the transcription start site for both gene sets. Although the positions of these symmetries are slightly different relative to the '−35' region of the two gene sets, the sizes are comparable (60 bp for PC, 74 bp for APC). The possible function of these dyad symmetries is not known but might be involved in gene regulation (Gilbert and Maxam, 1973). Isolation of these same genes from other eukaryotic algae and/or isolation of structural genes for other components of the phycobilisome in *C. paradoxa* may help in understanding the relevance and possible function of these structures.

Materials and methods

Materials

All chemicals used were of reagent grade. Alkaline phosphatase, DNA polymerase I, DNA polymerase I Klenow fragment and S1 nuclease were obtained from Boehringer Mannheim; polynucleotide kinase and T4 DNA ligase from Bethesda Research Laboratories; deoxyribonuclease I from Sigma; and restriction enzymes from Bethesda Research Laboratories and New England Biolabs. [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37GBq) and [α-^{32}P]dCTP (800 Ci/mmol) were obtained from Amersham.

Vectors, strains and growth conditions

The pBR322-derivative pUC8 (Vieira and Messing, 1982) was used for subcloning by transforming ligated plasmids into *E. coli* strain JM83 (Messing, 1979). *C. paradoxa*, a subculture of UTEX LB555, was grown as described (Lemaux and Grossman, 1984).

Isolation of plastid DNA from C. paradoxa and subcloning of plastid fragments

Plastid DNA was isolated from *C. paradoxa* as described (Lemaux and Grosman, 1984). DNA fragments to be subcloned were generated by digestion of plastid DNA with restriction enzymes and subsequent isolation from agarose gels [1% in Tris-acetate EDTA (TEA) or Tris-borate EDTA (TBE) buffer (Maniatis *et al.*, 1982)]. Fragments, electroeluted from agarose gels, were purified by passage over NACS52 resin (according to specifications of Bethesda Research Laboratories) and then ligated (manufacturer's specifications) into pUC8 vector which had been treated with the appropriate restriction endonuclease and alkaline phosphatase (manufacturer's specifications) and isolated from preparative polyacrylamide gels (TBE buffer). Ligated DNA was transformed into competent (Bedbrook, 1982) JM83 and resultant colonies containing inserts (lac$^-$ phenotype) were screened by restriction digestion and Southern hybridization for the presence of the appropriate sequences.

Restriction mapping and Southern hybridizations

Restriction mapping of *E. coli* clones was done by classical restriction analysis (Southern, 1979) of total plasmid DNA and isolated insert. In some cases fragments were labeled with polynucleotide kinase (Maxam and Gilbert, 1980) or filled in with DNA polymerase Klenow fragment (Sanger *et al.*, 1977) to aid in analysis. To determine the location of the APC genes on the plastid genome, appropriate DNA fragments were labeled by nick-translation (Rigby *et al.*, 1977) and hybridized to plastid DNA digested with various restriction enzymes. Southern hybridizations (12 − 18 h at the temperatures indicated in the figure legends) were carried out, without pre-hybridization, in 0.5 M NaCl/0.1 M NaH_2PO_4/0.1 M Tris base/0.002 M EDTA/0.1% SDS using labeled DNA that had been fragmented by boiling for 5 min in 0.5 M NaOH. After hybridization the filters were washed three times (10 min each) in 0.01 M PO_4 buffer pH 7.0/0.002 M EDTA/0.1% SDS at room temperature and, where indicated, 30 min in 0.05 M PO_4 buffer, pH 7.0/0.002 M EDTA/0.1% SDS at 42°C. The washed filters were exposed to Kodak XAR-5 film at −80°C for the times indicated.

Sequencing

Fragments for sequence analysis were generated by digesting with restriction endonucleases and (i) treating with alkaline phosphatase, labeling with [γ-^{32}P]ATP and polynucleotide kinase and recutting or strand-separating (Maxam and Gilbert, 1980) and/or (ii) using DNA polymerase Klenow fragment to fill-in protruding ends with the appropriate [α-^{32}P]dNTP and recutting or strand-separating. Labeled fragments recut with restriction endonucleases or treated with DMSO for strand-separation were loaded onto preparative polyacrylamide gels (TBE buffer), visualized by autoradiography, and electroeluted and purified by miltiple NH_4OAc precipitations. Fragments were sequenced by the Maxam and Gilbert method (1980) using five different overlapping chemistries and read from multiple loadings on 8% and 20% sequencing gels.

Isolation of poly(A)$^-$ RNA and Northern analysis

RNA isolations (Cashmore *et al.*, 1978) were performed on logarithmically grown cultures, which were pelleted and lysed by suspension in 0.5 M Tris-HCl, pH 9.0/0.015 M EDTA/1.0% SDS/0.005 M dithiothreitol. The lysate was extracted sequentially with one volume phenol, one volume phenol:chloroform (1:1 v/v) and twice with chloroform. The aqueous phase was precipitated with 0.8 volumes of isopropanol and further purified by precipitation with 2 M LiCl and twice with sodium chloride and ethanol. Poly(A)$^-$ RNA was selected by passage over poly(U)-derivatized Sepharose 4B (Pharmacia) (Lindberg and Perrson, 1972) and the fraction not bound to the derivatized Sepharose was used as poly(A)$^-$ RNA. The poly(A)$^-$ RNA was treated with deionized formamide and formaldehyde and loaded onto a 1.5% agarose gel (MEN buffer) (Maniatis *et al.*, 1982). DNA standards for mol. wt. determinations were treated in the same manner as the RNA, and loaded onto a portion of the gel which was stained and photographed. DNA fragments on the remainder of the gel were transferred to nitrocellulose as described (Thomas, 1980) and baked for 3 h at 80°C *in vacuo*. Hybridizations with alkali-treated, nick-translated probes and washings were as described for Southern hybridizations.

S1 nuclease mapping

The 5′ end of the transcript for the PC genes was determined by S1 nuclease digestion using a 450-bp fragment encoding the amino-terminal 17 amino acids of β-PC. To prepare this fragment, a plasmid, pCPC4067 (Lemaux and Grossman, 1984), was digested with *Eco*RI, treated with alkaline phosphatase and polynucleotide kinase and then recut with *Hind*III generating a 450-bp fragment labeled at its 5′ end. The 5′ end of the transcript for the APC genes was determined using a 165-bp fragment which encodes the amino-terminal 18 amino acids of α-APC. This was generated by digesting pCPC1422 with *Dde*I (Figure 2B), treating it with alkaline phosphatase and polynucleotide kinase and redigesting it with *Eco*RI. The 165-bp fragment labeled at its 5′ end contained the amino-terminal 18 amino acids of α-APC. The S1 nuclease digestions were carried out with slight modifications according to published procedures (Berk and Sharp, 1977; Maniatis *et al.*, 1982; Weaver and Weissman, 1979). Briefly, poly(A)$^-$ RNA from *C. paradoxa* (7.5 μg for PC reaction; 120 μg for APC) was ethanol-precipitated with the appropriate labeled DNA fragment (0.05 pmol [^{32}P]PC DNA; 0.01 pmol [^{32}P]APC DNA). (Carrier tRNA was added to the PC reaction to give a final RNA concentration during hybridization of 1 μg/μl.) The pellet was dried, resuspended in 30 μl of hybridization buffer [0.04 M Pipes buffer, pH 6.4/0.4 M NaCl/0.001 M EDTA in 80% (v/v) deionized formamide] and denatured at 85°C for 15 min. Hybridizations were carried out overnight at 34°C and 37°C for the PC transcript and 35°C and 38°C for the APC transcript. S1 digestion buffer (0.25 M NaCl/0.03 M NaOAc, pH 4.8/0.001 M $ZnCl_2$) and S1 nuclease (10^4 U/ml for PC; 3 x 10^3 U/ml for APC) were added for 60 min at 30°C. Reactions were terminated by adding 0.5 M NH_4OAc/0.01 M EDTA/0.5 μg/ml tRNA and 50% isopropanol (final concentrations). Pelleted nucleic acids were then prepared as for sequencing and loaded with the appropriate sequencing chemistries on 8% sequencing gels.

Acknowledgements

The authors wish to thank Glenn Ford for his expert help with graphic illustration and nucleotide sequence analysis, Sabrina Bennahimias and Linda Morris for excellent technical assistance, Dr. Pam Conley, Laura Green, Eugenio DeHostos and Dr. Terri Lomax for helpful discussions, Mark Dudley for help with the Northern analyses and Loretta Tayabas for excellent secretarial assistance. PGL was supported by a McKnight Foundation postdoctoral fellowship, a Carnegie

Institution of Washington fellowship, and by a National Institutes of Health grant (GM 33436-01). ARG was supported by Carnegie Institution of Washington and National Institutes of Health grant (GM 33436-01).

References

Arntzen,C.J. (1978) *Curr. Top. Bioenerg.*, **8**, 111-160.

Bedbrook,J.R. (1982) in Edelman,M., Hallick,R.B. and Chua,N.-H. (eds.) *Methods in Chloroplast Molecular Biology*, Elsevier, Amsterdam, The Netherlands, pp. 521-530.

Belasco,J.G., Beatty,J.T., Adams,C.W., von Gabain,A. and Cohen,S.N. (1985) *Cell*, **40**, 171-181.

Belford,H.S., Offner,G.D. and Troxler,R.F. (1983) *J. Biol. Chem.*, **258**, 4503-4510.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.

Bogorad,L. (1975) *Ann. Rev. Plant Physiol.*, **26**, 369-401.

Bohnert,H.J., Michalowski,C., Koller,B., Delius,H., Mucke,H. and Löffelhardt,W. (1983) in Schenk,H.E.A. and Schwemmler,W. (eds.) *Endocytobiology II*, deGruyter, Berlin, FRG, pp. 433-448.

Borbély,G. and Simoncsits,A. (1981) *Biochem. Biophys. Res. Commun.*, **101**, 846-852.

Bryant,D.A., Hixson,C.S. and Glazer,A.N. (1978) *J. Biol. Chem.*, **253**, 220-225.

Bryant,D.A. and Cohen-Bazire,G. (1981) *Europ. J. Biochem.*, **119**, 415-424.

Cashmore,A.R., Broadhurst,M.K. and Gray,R.E. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 655-659.

Conley,P.B., Lemaux,P.G. and Grossman,A.R. (1985) *Science*, in press.

Craig,I.W. and Carr,N.G. (1968) *Biochem. J.*, **106**, 361-366.

de Lorimier,R., Bryant,D.A., Porter,R.D., Liu,W.-Y., Jay,E. and Stevens,S.E.,Jr. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7946-7950.

DeLange,R.J., Williams,L.C. and Glazer,A.N. (1981) *J. Biol. Chem.*, **256**, 9558-9666.

Egelhoff,T. and Grossman,A.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3339-3343.

Fisher,R., Tuli,R. and Haselkorn,R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3393-3397.

Frank,G., Sidler,W., Widmer,H. and Zuber,H. (1978) *Hoppe-Seyler's Z. Pysiol. Chem.*, **259**, 1491-1507.

Freidenreich,P., Apell,G.S. and Glazer,A.N. (1978) *J. Biol. Chem.*, **253**, 212-219.

Gantt,E. (1981) *Annu. Rev. Plant Physiol.*, **32**, 327-347.

Gantt,E. and Lipschultz,C.A. (1974) *Biochemistry (Wash.)*, **13**, 2961-2966.

Gilbert,W. and Maxam,A. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 3581-3585.

Glazer,A.N. (1977) *Mol. Cell. Biochem.*, **18**, 125-140.

Glazer,A.N. (1982) *Annu. Rev. Microbiol.*, **36**, 171-196.

Glazer,A.N., Lundell,D.J., Yamanaka,G. and Williams,R.C. (1983) *Ann. Microbiol. (Inst. Pasteur)*, **134B**, 159-180.

Gold,L., Pribnow,D., Schneider,T., Shinedling,S., Singer,B.S. and Stormo,G. (1981) *Annu. Rev. Microbiol.*, **35**, 365-403.

Gray,M.W. and Doolittle,W.F. (1982) *Microbiol. Rev.*, **46**, 1-42.

Grossman,A.R., Talbot,L. and Egelhoff,T. (1983) *Carnegie Inst. Washington Yearbook*, **82**, 112-116.

Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, **11**, 2237-2255.

Hiller,R.G. and Goodchild,D.J. (1981) in Hatch,M.D. and Boardman,D.J. (eds.) *Biochemistry of Plants*, Vol. **8**, Academic Press, NY, pp. 1-49.

Köller,L.P., Wehrmeyer,W. and Schneider,H. (1977) *Arch. Microbiol.*, **112**, 61-67.

Kuntz,M., Crouse,E.J., Mubumbila,M., Burkard,G., Weil,J.-H., Bohnert,H.J., Mucke,H. and Löffelhardt,W. (1984) *Mol. Gen. Genet.*, **194**, 508-512.

Lemaux,P.G. and Grossman,A.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4100-4104.

Lindberg,U. and Perrson,R. (1972) *Eur. J. Biochem.*, **31**, 246-254.

Lundell,D.J., Williams,R.C. and Glazer,A.N. (1981) *J. Biol. Chem.*, **256**, 3580-3592.

Maniatis,R., Fritsch,E.F. and Sambrook,J. (1982) in *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Margulis,L. (1970) *Origin of Eukaryotic Cells*, published by Yale University Press, New Haven, CT.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.

Messing,J. (1979) *Recombinant DNA Technical Bulletin*, NIH Publication No. 79-99.2, No. 2, 43-48.

Movva,N.R., Nakamura,K. and Inouye,M. (1980) *J. Mol. Biol.*, **43**, 317-328.

Offner,G.D. and Troxler,R.F. (1983) *J. Biol. Chem.*, **258**, 9931-3340.

Offner,G.D., Brown-Mason,A.S., Ehrhardt,M.M. and Troxler,R.F. (1981) *J. Biol. Chem.*, **256**, 12167-12175.

Pilot,T.J. and Fox,J.L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6983-6987.

Ramus,J., Beale,S.I., Mauzerall,D. and Howard,K.L. (1976) *Mar. Biol.*, **32**, 223-229.

Redlinger,T. and Gantt,E. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5542-5546.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.

Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319-353.

Sangar,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Schreier,P.H. and Bohnert,H.J. (1985) in Schell,J.S. and Starlinger,P. (eds.) *The Impact of Gene Transfer Techniques in Eukaryotic Cell Biology*, Springer-Verlag, Berlin, Heidelberg, FRG, pp. 91-102.

Schwarz,Z. and Kossel,H. (1979) *Nature*, **279**, 520-522.

Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.

Siegelman,H.W. and Kycia,J.H. (1982) *Plant Physiol.*, **79**, 887-897.

Southern,E. (1979) *Methods Enzymol.*, **68**, 152-176.

Tanaka,N., Cramer,J.H. and Rownd,R.H. (1976) *J. Bacteriol.*, **127**, 619-636.

Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5203.

Tohdoh,N. and Sugiura,M. (1982) *Gene*, **17**, 213-218.

Trench,R.K. and Ronzio,G.S. (1978) *Proc. R. Soc. Lond., Ser. B*, **202**, 445-462.

Troxler,R.F., Foster,J.A., Brown,A.S. and Franzblau,C. (1975) *Biochemistry (Wash.)*, **14**, 268-274.

Troxler,R.F., Ehrhardt,M.M., Brown-Mason,A.S. and Offner,G.D. (1981) *J. Biol. Chem.*, **256**, 12176-12184.

Vieira,J. and Messing,J. (1982) *Gene*, **9**, 259-268.

Wang,R.T. and Myers,J. (1976) *Photochem. Photobiol.*, **23**, 411-414.

Wang,R.T., Stevens,C.L.R. and Myers,J. (1977) *Photochem. Photobiol.*, **25**, 103-108.

Weaver,R.F. and Weissman,C. (1979) *Nucleic Acids Res.*, **5**, 1175-1193.

Whitfeld,P.R. and Bottomley,W. (1983) *Annu. Rev. Plant Physiol.*, **34**, 279-310.

Williams,V.P. and Glazer,A.N. (1978) *J. Biol. Chem.*, **253**, 202-211.

Zuber,H. (1983) in Papageorgiou,G. and Packer,L. (eds.) *Photosynthetic Prokaryotes: Cell Differentiation and Function*, Elsevier, NY, pp. 23-42.

Zurawski,G., Bottomley,W. and Whitfeld,P.R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6260-6264.

Received on 7 May 1985

Note added in proof

Because of complications with gene designations in *E. coli*, the nomenclature for the phycobiliprotein genes must be revised to: α-phycocyanin, *pcy*A; β-phycocyanin, *pcy*B; α-allophycocyanin, *apc*A and β-allophycocyanin, *apc*B.