JOURNAL OF BACTERIOLOGY, Dec. 1988, p. 5512–5521
0021-9193/88/125512-10$02.00/0

Vol. 170, No. 12

Genes Encoding Core Components of the Phycobilisome in the Cyanobacterium *Calothrix* sp. Strain PCC 7601: Occurrence of a Multigene Family

JEAN HOUMARD,* VÉRONIQUE CAPUANO, THÉRÈSE COURSIN, AND NICOLE TANDEAU DE MARSAC

Unité de Physiologie Microbienne, Centre National de la Recherche Scientifique URA150, Département de Biochimie et Génétique Moléculaire, Institut Pasteur, 28 rue du Dr Roux, F-75724 Paris Cedex 15, France

Received 29 April 1988/Accepted 9 September 1988

The phycobilisome is the major light-harvesting complex of cyanobacteria. It is composed of a central core from which six rods radiate. Allophycocyanin, an αβ oligomer (α^{AP} and β^{AP}), is the main component of the core which also contains three other phycobiliproteins (α^{APB}, $\beta^{18.3}$, and L_{CM}^{92}) and a small linker polypeptide ($L_C^{7.8}$). By heterologous DNA hybridization, two *Eco*RI DNA fragments of 3.5 and 3.7 kilobases have been cloned from the chromatically adapting cyanobacterium *Calothrix* sp. strain PCC 7601. Nucleotide sequence determination has allowed the identification of five *apc* genes: *apcA1* (α^{AP1}), *apcA2* (α^{AP2}), *apcB1* (β^{AP1}), *apcC* ($L_C^{7.8}$), and *apcE* (L_{CM}^{92}). Four of these genes are adjacent on the chromosome and form the *apcEA1B1C* gene cluster. In contrast, no genes have been found close to the *apcA2* gene which is carried by the 3.5-kilobase *Eco*RI fragment. Transcriptional analysis and 5′-end-mapping experiments were performed. The results obtained demonstrate that the *apcEA1B1C* gene cluster forms an operon from which segmented transcripts originate, whereas the *apcA2* gene behaves as a monocistronic unit. Qualitatively, the same transcripts were identified regardless of the light wavelengths received during cell growth. The deduced amino acid sequences of the *apc* gene products are very similar to their known homologs of either cyanobacterial or eucaryotic origin. It was interesting, however, that in the *apcA1* and *apcA2* genes, whose products correspond to α-type allophycocyanin subunits, nucleotide sequences were more conserved (67%) than were the deduced amino acid sequences (59%).

Cyanobacteria constitute a very large group of procaryotes that perform O_2-evolving photosynthesis, a function characteristic of eucaryotic plants. Their light-harvesting apparatus is composed of protein-chlorophyll *a* complexes and of highly ordered multimolecular structures, the phycobilisomes (PBsomes) (39). Most cyanobacteria possess hemidiscoidal PBsomes that are attached to the surface of the photosynthetic membranes and are made up of at least 12 different polypeptide chains. They can be easily purified as water-soluble complexes (18). They generally consist of two structural domains, a central core from which six rods radiate (5). Rods are composed of stacked disks, each of which represents a hexameric complex of specific phycobiliprotein (PB) and linker polypeptide molecules. Phycocyanin (PC) hexamers are always present, but in some strains the PBsomes also contain either phycoerythrin or phycoerythrocyanin disks in the distal portions of the rods. Most of the cyanobacterial PBsomes have a three-cylinder core, but in a few strains the core may contain only two cylinders. Core substructure has been analyzed in detail for two strains: *Synechocystis* sp. strain PCC 6701 (tricylindrical core) and *Synechococcus* sp. strain PCC 6301 (bicylindrical core) (for reviews, see references 17, 20, and 21). Allophycocyanin (AP; α^{AP} and β^{AP}) is the major constituent of the core which also contains three minor PB species: a β-type AP subunit ($\beta^{18.3}$), the APB (α^{APB}), and the so-called anchor polypeptide (L_{CM}^{75-120}). (PBsome components are abbreviated according to the nomenclature proposed by Glazer [20]; the corresponding genes are abbreviated as proposed by J. Houmard and N. Tandeau de Marsac [Methods Enzymol., in press]. The subscripts in the linker polypeptides denote their location [C, core; R, rod; CM, core-membrane], and the superscripts indicate their molecular masses.) In addition, a small linker polypeptide ($L_C^{7.8}$) completes the structure. Within the core, AP subunits are grouped into trimers, either alone or in conjunction with the various other minor components, with four trimers being stacked to form a cylinder. In the tricylindrical cores, two basal cylinders interface with the thylakoid membrane, with the third cylinder on top. A model for this asymmetrical core substructure has recently been proposed (1, 19). In addition to biochemical and biophysical studies, molecular biology techniques have been used to complete the structural and functional analysis of PBsomes. Genes encoding some of the PBsome components have already been isolated from various cyanobacteria (for a review, see Houmard and Tandeau de Marsac, in press).

Among cyanobacteria, *Calothrix* sp. strain PCC 7601 is a strain which can adapt the pigment composition of its PBsomes to the incident light wavelengths, a phenomenon known as complementary chromatic adaptation (for a review, see reference 40). Recently, all of the genes which encode rod components, except the linker polypeptides associated with phycoerythrin, have been isolated from *Calothrix* sp. strain 7601 (6, 9, 10, 26, 28, 29). It has also been demonstrated that transcriptional regulation occurs during complementary chromatic adaptation. Under green-light conditions, genes encoding phycoerythrin (*cpeBA*) are turned on whereas those encoding PC2 (*cpcB2A2*) and its associated linker polypeptides (*cpcH2I2D2*) are turned off. Under red light, the opposite situation is observed (9, 10, 26, 28; N. Tandeau de Marsac, D. Mazel, T. Damerval, G. Guglielmi, V. Capuano, and J. Houmard, Photosynth. Res., in press). On the other hand, two sets of PC genes (*cpcB1A1* and *cpcB3A3*) are expressed regardless of the light wave-

* Corresponding author.

length received by the cells during growth (29). Besides this kind of regulation, light-independent regulations, at both transcriptional and posttranscriptional levels, have also been recently documented for genes encoding rod components (Tandeau de Marsac et al., in press). To extend our studies on the photoregulation of gene expression in *Calothrix* sp. strain 7601, we have undertaken the characterization of the genes that encode the components of the tricylindrical cores of its PBsomes. The characterization of the *apcD* gene, which corresponds to α^{APB} (24), and a partial nucleotide sequence of the *apcA1* (α^{AP1}) and *apcB1* (β^{AP1}) genes (11) have been previously reported. In this study we present the complete nucleotide sequence and a transcriptional analysis of the genes which encode α^{AP1}, β^{AP1}, and $L_C^{7.8}$ (the *apcA1B1C* gene cluster) and of the gene *apcA2*, which corresponds to an α-type AP subunit that has not yet been biochemically characterized.

MATERIALS AND METHODS

Materials. All chemicals were of reagent grade. Restriction enzymes, calf intestinal alkaline phosphatase, T4 polynucleotide kinase, S1 nuclease, and T4 DNA ligase were purchased from either Boehringer (F-38242; Meylan, France) or Genofit (CH-1212; Geneva). Klenow fragment of DNA polymerase I, [γ-^{32}P]ATP (110 TBq/mmol), [α-^{32}P]dXTP (110 TBq/mmol), [α-^{35}S]dATP (15 TBq/mmol), Hybond-N membranes for hybridization experiments, and a nick translation kit were from Amersham (F-91944; Les Ulis, France). The Cyclone System was from International Biotechnologies, Inc. (IBI; distributed by Genofit). Nucleotides, dideoxynucleotide triphosphates, M13 derivatives, pTZ18R, and pTZ19R were purchased from PL-Pharmacia (F-91943; Les Ulis, France). Enzymes were used according to the specifications of the suppliers. Acrylamide, bisacrylamide, and other reagents for the preparation of polyacrylamide gels used for DNA sequencing were obtained from BDH Chemicals, Ltd. (distributed by Interchim; F-75013; Paris). Nitrocellulose membranes were purchased from Schleicher & Schuell (distributed by Cera-Labo; F-93307; Aubervilliers).

Culture conditions and DNA and RNA isolations. Culture conditions for *Calothrix* sp. strain PCC 7601 (*Fremyella diplosiphon* UTEX 481), as well as chromatic illuminations and total DNA and total RNA purifications have been described previously (28). Small-scale plasmid extractions were performed on overnight-grown cultures of *Escherichia coli* as described earlier (23). Large-scale plasmid preparation, DNA fragment isolation, and agarose gel electrophoresis were performed as described by Maniatis et al. (27).

Genomic library construction and screening. Construction of the *Calothrix* strain 7601 genomic library into λEMBL3 has been reported previously (41). It results from ligation of a partial *Sau*3A digest into the *Bam*HI site of λEMBL3. The partial DNA library was constructed by ligation of *Eco*RI DNA fragments (3.5 to 4 kilobases [kb]) into the *Eco*RI site of pUC8 as described elsewhere (29). Standard methods were used for screening by plaque hybridization and in situ colony hybridization (27).

Hybridization analysis. Nick translation and Southern hybridization experiments were performed as described previously (41). Heterologous DNA hybridizations were done at 60°C for 16 to 24 h without carrier DNA in the prehybridization and hybridization solutions. Filter washings were performed at room temperature. Northern (RNA) transfers and hybridization were performed as described by Damerval et al. (13) at 55°C in 50% formamide.

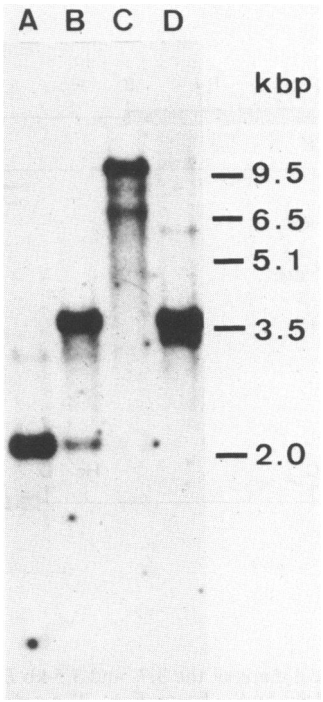

FIG. 1. Fluoro-autoradiogram of a Southern blot probed with a 0.8-kb *Eco*RI-*Pst*I fragment carrying part of the *apcAB* operon from *Synechococcus* sp. strain PCC 6301. Restriction endonuclease digests of *Calothrix* sp. strain 7601 total DNA were electrophoresed, transferred to nitrocellulose membranes, and hybridized at 60°C. Lanes: A, *Hin*cII; B, *Hin*dIII-*Eco*RI; C, *Hin*dIII; D, *Eco*RI. The bars at the right show the positions of size markers generated by digestion of λ DNA with *Hin*dIII and *Hin*dIII-*Eco*RI.

DNA sequence analysis. Large DNA fragments were subcloned into either pTZ18R or pTZ19R. Overlapping clones were obtained by using the Cyclone System from IBI adapted for single-stranded DNAs of the pTZ18R subclones (29). DNA sequence analysis was performed by the chain-termination method of Sanger et al. (37) on single-stranded DNA templates according to the protocol recommended by Amersham. Computer analysis of the DNA sequence information obtained was done by using the programs developed by the Unité d'Informatique Scientifique of the Institut Pasteur (F-75015; Paris).

Mapping of the transcripts. End-labeled restriction fragments were generated with [γ-^{32}P]ATP and T4 polynucleotide kinase (27). The fragments were purified on either 4 or 20% polyacrylamide gels, depending on the length of the DNA fragment. S1 nuclease mappings and primer extension experiments were performed as described previously (24).

RESULTS

Gene isolation. Figure 1 shows the results obtained after hybridization of a Southern blot carrying restriction endonuclease digests of *Calothrix* strain 7601 total DNA with a 0.8-kb *Eco*RI-*Pst*I fragment from *Synechococcus* sp. strain PCC 6301. This probe carried all of the *apcA* and part of the *apcB* genes (25). Three hybridization bands were detected in both *Eco*RI and *Hin*dIII digests under conditions which allowed approximately 40% mismatching (7). Screening of the λEMBL3 and pUC8 libraries of *Calothrix* strain 7601 with the same probe allowed the isolation of the three *Eco*RI fragments. The 6.3-kb *Eco*RI fragment, which corresponded

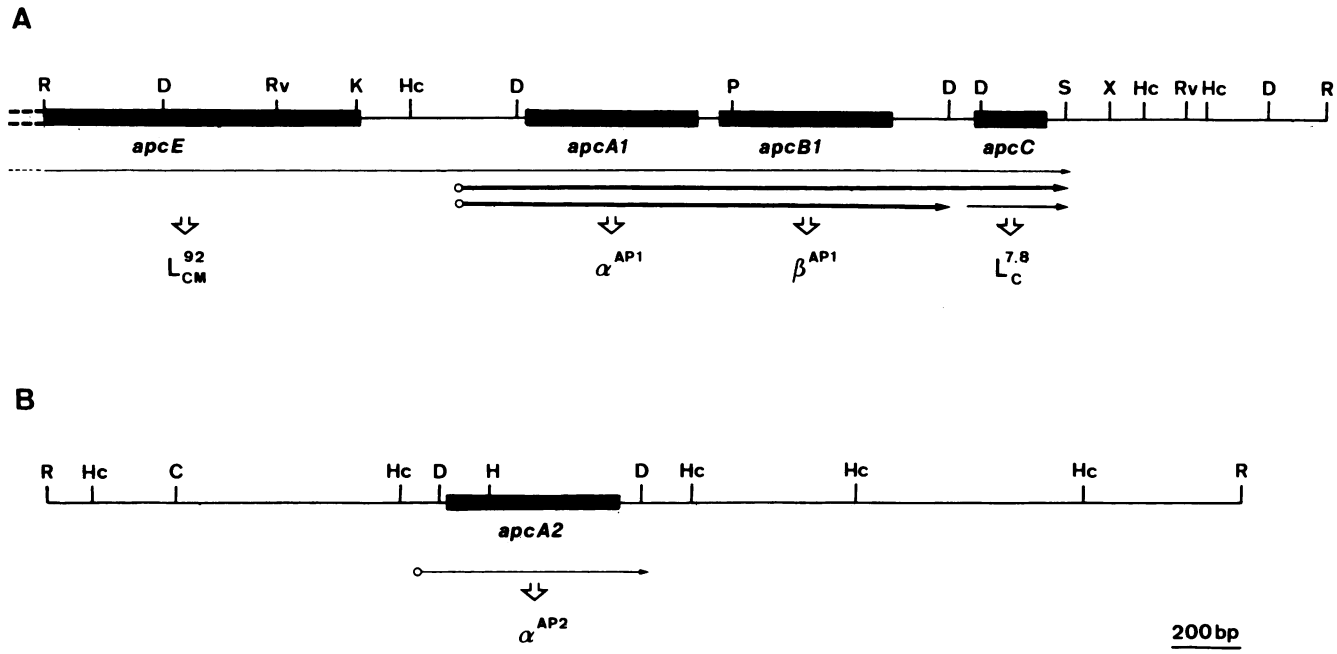

FIG. 2. Physical maps of the 3.7- and 3.5-kb *Eco*RI DNA fragments cloned in plasmid pPM29 (A) and pPM30 (B), respectively. Coding regions are indicated by the heavy lines. The gene designations and the encoded gene products are given below. Abbreviations: C, *Cla*I; D, *Dra*I; H, *Hin*dIII; Hc, *Hin*cII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; Rv, *Eco*RV; S, *Stu*I; X, *Xba*I. The horizontal arrows represent transcripts originating from these DNA fragments (the size of the *apcA2* transcript has not been established with certainty). The open circles at the extremities of the arrows indicate that the 5′ ends of the corresponding transcripts have been mapped. Thicker arrows correspond to the most abundant transcripts.

to the less-intense hybridization signal, carried the *apcD* gene, which encodes APB (α^{APB}) (24). A recombinant phage (λD4) carried the 3.5-kb DNA fragment, and a recombinant plasmid (pPM29) carried the 3.7-kb fragment. The 3.5-kb *Eco*RI fragment was subcloned from λD4 into the *Eco*RI site of plasmid vector pUC8 to generate plasmid pPM30. Restriction analysis and hybridization experiments were used to develop the physical maps of the 3.5- and 3.7-kb *Eco*RI DNA fragments (Fig. 2).

Nucleotide sequence analysis. The nucleotide sequences presented in Fig. 3 were completely determined on both strands. In both sequences *apcA*-like genes, designated *apcA1* and *apcA2*, were found. They both consisted of 162 codons. Sequence homologies with genes that encode PB-some components also allowed us to locate the *apcB1* gene (163 codons) 65 base pairs (bp) downstream from *apcA1* and the *apcC* gene (69 codons) 240 bp downstream from *apcB1* (Fig. 3). In contrast, no *apcB*- or *apcC*-like genes were found within 1-kb sequences on either side of the *apcA2* gene (data not shown). For each of these four genes, a sequence 5′ AGGA 3′ occurs just upstream (8 to 9 nucleotides [nt]) from the translational start codons. It is likely that these polypurine stretches may act as ribosome-binding sites, although the 3′ end of the *Calothrix* strain 7601 16S rRNA has not yet been determined.

Hypothetical palindromic sequences could be drawn within the DNA sequences presented in Fig. 3. Their thermodynamic stabilities were calculated according to the method of Salser (36) as modified by Cech et al. (8). In front of the *apcA1* coding region, a stem and loop structure (ΔG = −16.1 kcal/mol [1 cal = 4.184 J]) could exist from nt 25 through nt 94, with a hairpin loop of 28 nt. Within the *apcA1* gene, nt 456 to 465 could be paired with nt 477 to 486 (ΔG = −10.3 kcal/mol). No significantly stable secondary structure seemed to be present either in the *apcA1-apcB1* intercistronic region or within the *apcB1* coding region. In contrast, a large and highly stable structure (nt 1409 to 1444 being paired with nt 1487 to 1522; ΔG = −46.9 kcal/mol) and a less-stable one (nt 1542 to 1555 paired with 1564 to 1577; ΔG = −7.7 kcal/mol) could be formed in the *apcB1-apcC* intercistronic region. Finally, a stem and loop structure (ΔG = −15.3 kcal/mol) could exist downstream from the *apcC* coding region which would pair both nt 1826 to 1836 with 1880 to 1890 and nt 1856 to 1864 with nt 1871 to 1879. Similarly, possible palindromic sequences were detected in the DNA fragment which carries *apcA2*. In front of the coding sequence, nt 44 to 52 could be paired with nt 74 to 82 (ΔG = −7.5 kcal/mol). Another structure, nt 120 to 131 paired with nt 192 to 203 (ΔG = −13.4 kcal/mol), would place the putative ribosome-binding site within the stem of the structure and the AUG within the 60-nt hairpin loop. Within the coding region, nt 254 to 265 and nt 317 to 328, as well as nt 340 to 350 and nt 394 to 404, could give rise to structures with thermodynamic stabilities of −13.7 and −14.6 kcal/mol, respectively, but both had large hairpin loops (51 and 43 nt, respectively). At the end of the gene, nt 665 to 678 could interact with nt 697 to 710, generating a stable hairpin structure (ΔG = −29.1 kcal/mol).

The deduced amino acid sequences are presented in Fig. 4. The predicted molecular masses are 17,345 daltons for α^{AP1} (*apcA1*), 17,249 daltons for β^{AP1} (*apcB1*), 7,880 daltons for $L_C^{7.8}$ (*apcC*), and 17,832 daltons for α^{AP2} (*apcA2*). The codon usage patterns for these four genes are listed in Table 1. Nineteen *Calothrix* strain 7601 genes have been sequenced up to now, among which nine encode PB subunits (see legend of Table 1 for references). Comparison of codon usage in the four genes presented in this paper with the other *Calothrix* strain 7601 genes already reported revealed some

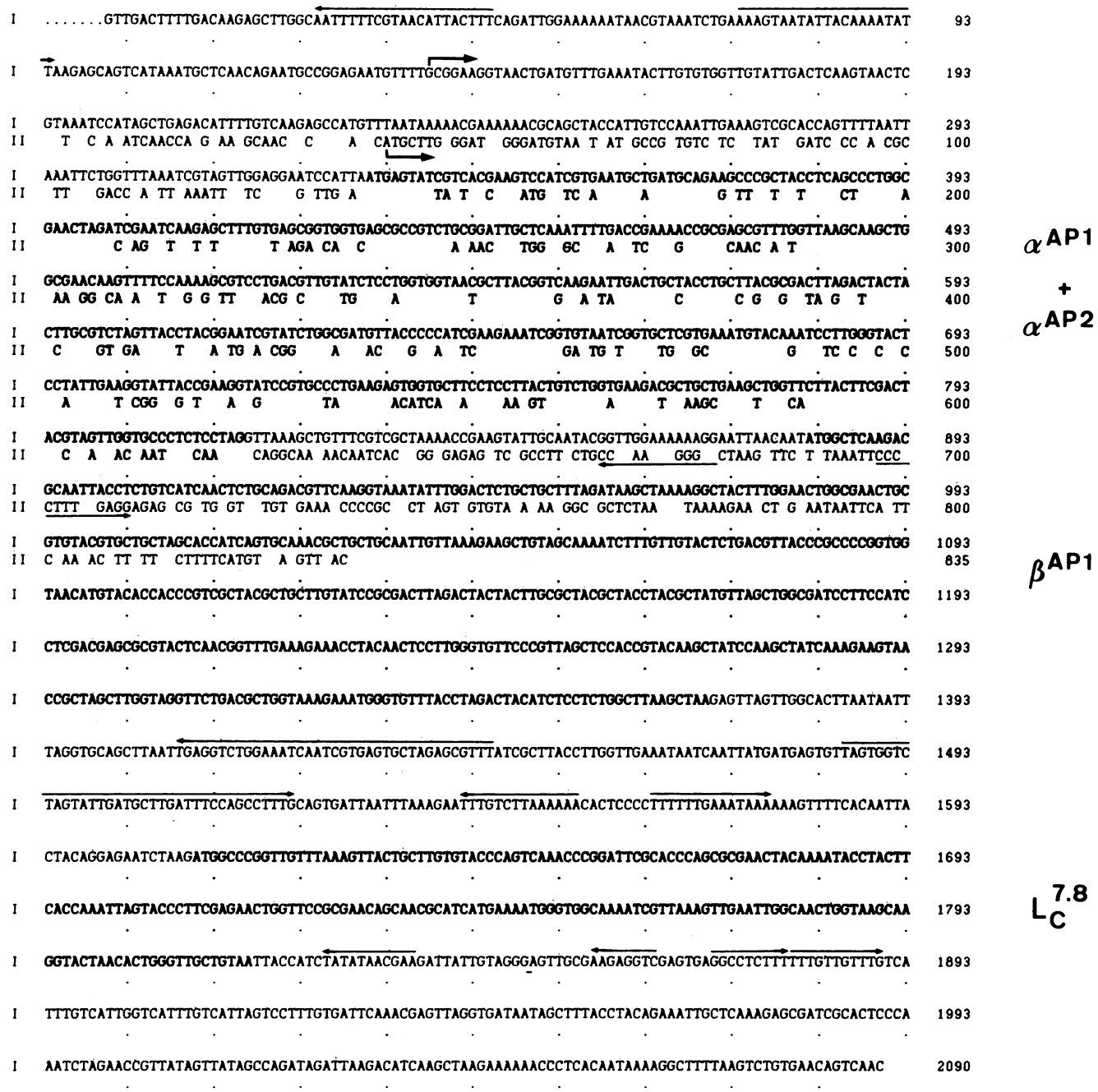

FIG. 3. Nucleotide sequences of the *Hin*cII fragments which carry the *apcA1B1C* (I) and the *apcA2* (II) genes. Numbering of nucleotides starts at the *Hin*cII site. The two sequences are compared after alignment of the sequences corresponding to the coding regions (in bold print). The encoded gene products are indicated in the right margin. The broken arrows indicate the mapped start sites of transcription. Palindromic sequences are indicated by arrows above (I) or below (II) the sequence.

differences. In both the *apcA1* and *apcB1* genes, the codon GAC is preferred over GAU for specifying aspartate, in contrast to the codon usage generally found. The same observation can be made for the codons which specify lysine residues in *apcA1*. A much higher degree of bias was found with the *apcA2* gene. In particular, the leucine codons UUA and UUG were used rarely, whereas CUA and CUU were used often; the UCC serine codon did not occur, but AGU was overemployed. Unusual codon bias was also observed for threonine, proline, lysine, and asparagine and to a lesser extent for glycine, valine, and isoleucine. In addition, as already observed for the *apcA* and *apcB* genes from *Synechococcus* sp. strain PCC 6301 (25), fewer synonymous codons are used for the *apcB1* gene than for the *apcA1* and *apcA2* genes.

Transcriptional analysis. Northern blot hybridizations were performed with various DNA probes. The *Dra*I DNA fragment, which only contains the *apcA1* and *apcB1* genes, revealed two major mRNA species of 1.4 and 1.7 kb; the *Dra*I-*Eco*RV fragment, which contains nearly all of the *apcC* gene plus 0.4 kb of its 3′-flanking sequence, also revealed the 1.7-kb transcript and a small transcript of about 0.3 kb (Fig.

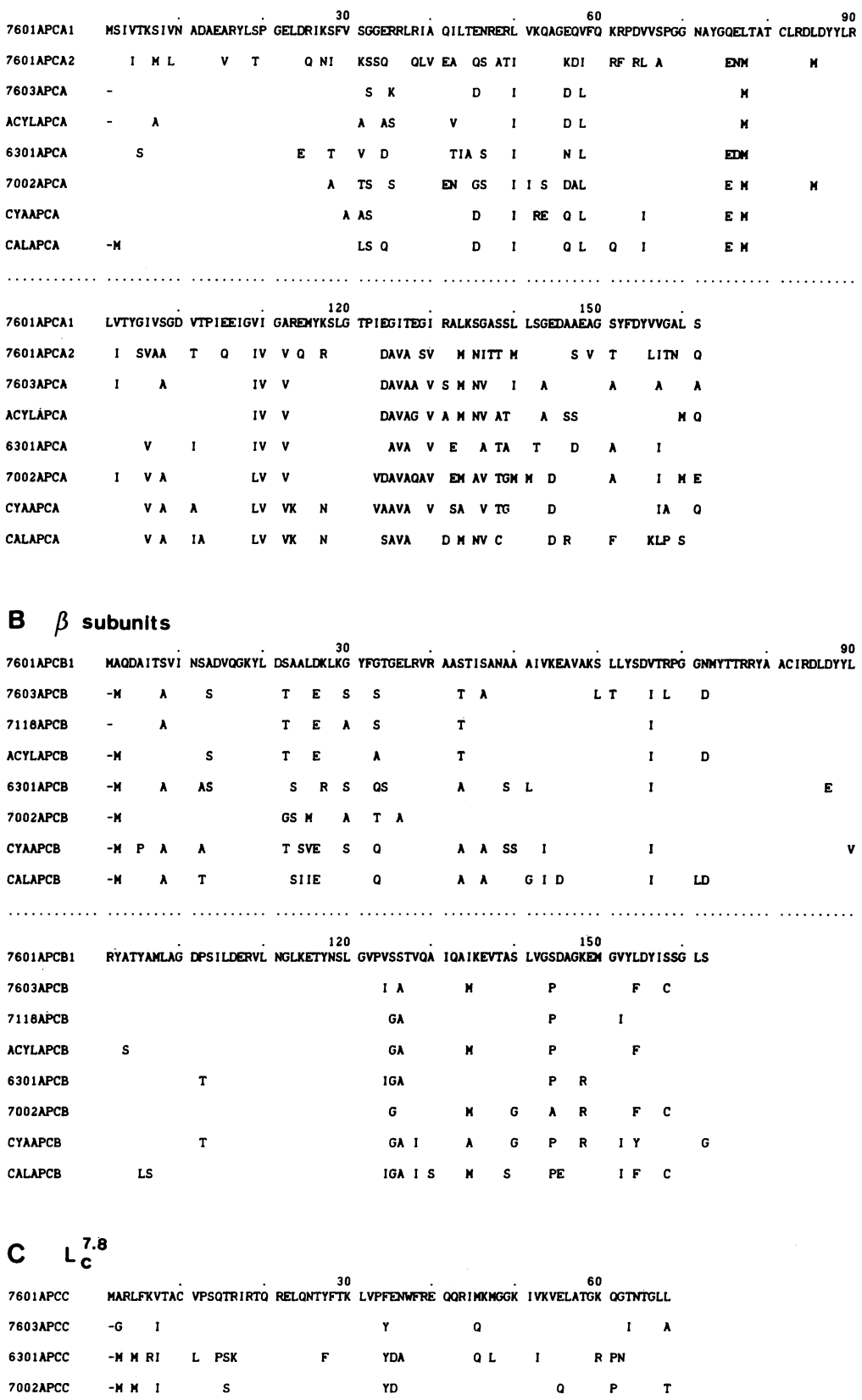

FIG. 4. Comparison of the amino acid sequences of α^{AP} (A), β^{AP} (B), and $L_C^{7.8}$ (C). Only residues which differ from those found at the equivalent position in the *Calothrix* sp. strain 7601 sequences of α^{AP1}, β^{AP1}, or $L_C^{7.8}$ are shown. The single-letter notation for amino acids is that recommended by the International Union of Pure and Applied Biochemistry-IUB. Organisms and references are as follows: 7601APCA1, APCA2, APCB1, and APCC refer to the deduced amino acid sequences presented in this paper; 7603APCA, APCB, and APCC are sequences determined for *M. laminosus* (*Fischerella* sp. strain PCC 7603) subunits (16, 38); 7118APCB is the amino acid sequence of β^{AP} from *Anabaena variabilis* (*Nostoc* sp., PCC 7118) (14); ACYLAPCA and APCB are the sequences determined for *A. cylindrica* subunits (30); 6301APCA, APCB, and APCC, as well as 7002APCA, APCB, and APCC, are the amino acid sequences deduced from the corresponding genes of *Synechococcus* sp. strain PCC 6301 (25) and PCC 7002 (D. A. Bryant, personal communication), respectively; CYAAPCA and APCB are deduced amino acid sequences from *Cyanophora paradoxa* (4); CCALAPCA and APCB are the amino acid sequences determined for the *C. caldarium* subunits (32). Dashes indicate missing residues.

TABLE 1. Codon usage in *apc* genes and in 19 *Calothrix* sp. strain 7601 genes

Residue and codon	Codon usage (no.)[a]				
	Calothrix strain PCC 7601	*apcA1*	*apcA2*	*apcB1*	*apcC*
Arg					
CGA	13 (6)	1 (8)	0	0	0
CGC	59 (27)	4 (33)	3 (27)	5 (62)	4 (67)
CGG	28 (13)	1 (8)	3 (27)	0	2 (33)
CGU	84 (39)	6 (50)	3 (27)	3 (37)	0
AGA	30 (14)	0	1 (9)	0	0
AGG	4 (2)	0	1 (9)	0	0
Leu					
CUA	48 (14)	2 (13)	4 (29)	2 (12)	1 (17)
CUC	23 (7)	2 (13)	3 (21)	2 (12)	0
CUG	41 (12)	3 (20)	2 (14)	1 (6)	1 (17)
CUU	29 (8)	0	2 (14)	0	0
UAA	106 (30)	3 (20)	1 (7)	4 (25)	1 (17)
UUG	101 (29)	5 (33)	2 (14)	7 (44)	3 (50)
Ser					
UCA	28 (10)	0	1 (10)	0	0
UCC	51 (18)	5 (36)	0	4 (25)	0
UCG	8 (3)	0	0	0	0
UCU	86 (31)	4 (29)	2 (20)	7 (44)	0
AGC	74 (26)	3 (21)	3 (30)	4 (25)	0
AGU	34 (12)	2 (14)	4 (40)	1 (6)	1 (100)
Thr					
ACA	51 (26)	0	0	0	0
ACC	86 (43)	5 (62)	4 (28)	9 (90)	4 (50)
ACG	5 (3)	1 (12)	1 (7)	0	0
ACU	57 (29)	2 (25)	9 (64)	1 (10)	4 (50)
Pro					
CCA	29 (21)	0	4 (80)	0	0
CCC	42 (30)	1 (20)	0	2 (67)	2 (100)
CCG	8 (6)	0	0	0	0
CCU	59 (43)	4 (80)	1 (20)	1 (33)	0
Ala					
GCA	162 (34)	1 (7)	4 (29)	5 (23)	1 (33)
GCC	42 (9)	3 (21)	1 (7)	0	1 (33)
GCG	35 (7)	0	2 (14)	0	0
GCU	239 (50)	10 (71)	7 (50)	17 (77)	1 (33)
Gly					
GGA	41 (16)	1 (6)	4 (36)	1 (8)	0
GGC	58 (22)	3 (17)	2 (18)	4 (31)	1 (20)
GGG	13 (5)	0	1 (9)	0	1 (20)
GGU	147 (57)	14 (78)	4 (36)	8 (61)	3 (60)
Val					
GUA	74 (30)	4 (31)	1 (9)	6 (46)	2 (40)
GUC	25 (10)	1 (8)	1 (9)	1 (8)	0
GUG	29 (12)	2 (15)	2 (18)	0	0
GUU	122 (49)	6 (46)	7 (64)	6 (46)	3 (60)
Ile					
AUA	13 (6)	0	1 (7)	0	0
AUC	95 (44)	8 (67)	3 (21)	7 (78)	2 (67)
AUU	109 (50)	4 (33)	10 (71)	2 (22)	1 (33)
Lys					
AAA	74 (60)	1 (17)	0	7 (87)	5 (83)
AAG	49 (40)	5 (83)	5 (100)	1 (12)	1 (17)
Asn					
AAC	93 (62)	2 (67)	1 (17)	5 (100)	2 (67)
AAU	57 (38)	1 (33)	5 (83)	0	1 (33)
Gln					
CAA	138 (75)	5 (100)	6 (67)	4 (100)	4 (67)
CAG	45 (25)	0	3 (33)	0	2 (33)
His					
CAC	16 (40)	0	0	0	0
CAU	24 (60)	0	0	0	0
Glu					
GAA	179 (87)	12 (86)	7 (87)	5 (83)	3 (75)
GAG	26 (13)	2 (14)	1 (12)	1 (17)	1 (25)
Asp					
GAC	55 (30)	5 (62)	3 (33)	9 (82)	0
GAU	131 (70)	3 (37)	6 (67)	2 (18)	0
Tyr					
UAC	88 (61)	8 (100)	7 (87)	11 (92)	1 (100)
UAU	56 (39)	0	1 (12)	1 (8)	0
Cys					
UGC	21 (70)	1 (100)	1 (100)	0	0
UGU	11 (30)	0	0	1 (100)	1 (100)
Phe					
UUC	56 (47)	2 (67)	1 (25)	0	3 (75)
UUU	63 (53)	1 (33)	3 (75)	1 (100)	1 (25)
Met, AUG	74	2	7	4	3
Trp, UGG	34	0	0	0	1
Ter					
UAA	8 (42)	0	0	1	1
UAG	8 (42)	1	1	0	0
UGA	3 (16)	0	0	0	0

[a] The percentage of synonymic use of a codon (in parentheses) was obtained by dividing the number of times that a codon was used by the number of times that all of the codons specifying the same amino acid were used. The codon usage for *Calothrix* sp. strain PCC 7601 was derived from a compilation of the 19 sequenced genes described thus far, i.e., 3,697 codons (6, 13, 24, 26, 28, 29, 31, 33, 41; Tandeau de Marsac et al., in press). Ter, Chain termination.

5). The *Pst*I-*Eco*RV fragment, which did not bear the *apcA1* gene, revealed the three transcripts. In addition, minor transcripts of about 5.6 and 2.8 kb were detected. These mRNA species are 30 to 50 times less abundant than the 1.4-kb transcript. On the other hand, no transcript was detected with the *Xba*I-*Eco*RV DNA fragment located downstream from the *apcC* gene (data not shown).

Nuclease S1 mapping and primer extension experiments were performed. *Dra*I-*Hin*cII DNA fragments of 303 and 112 bp were used for S1 nuclease analysis of the transcription initiation site for the *apcA1B1C* operon and the *apcA2* gene, respectively. A 51-bp *Dra*I-*Alu*I primer corresponding to the *apcA1B1C* 5′-flanking sequence and a 26-bp *Dra*I-*Pvu*I fragment corresponding to the *apcA2* 5′-flanking sequence were extended after annealing with total RNA. All of these experiments were performed with RNA preparations obtained from cells grown under both red and green light. For

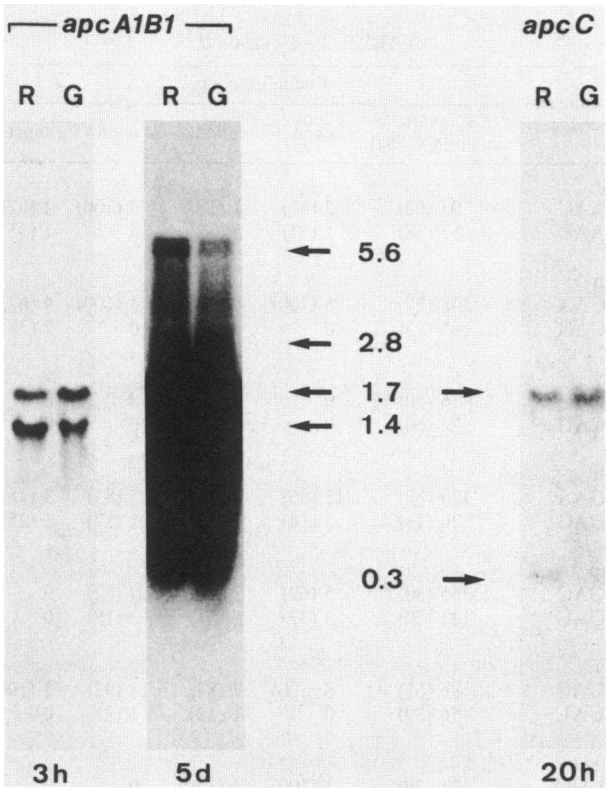

FIG. 5. Northern analysis of the mRNAs transcribed from the *apcA1B1C* gene cluster. Total RNAs extracted from *Calothrix* sp. strain 7601 cells grown under either red (R) or green (G) illumination were hybridized at 55°C in 50% formamide with an *apcA1B1* probe (*Dra*I DNA fragment [Fig. 2A]) or an *apcC* probe (*Dra*I-*Eco*RV DNA fragment [Fig. 2A]). Sizes are indicated in kilobases. The exposure times are mentioned below the fluoro-autoradiograms.

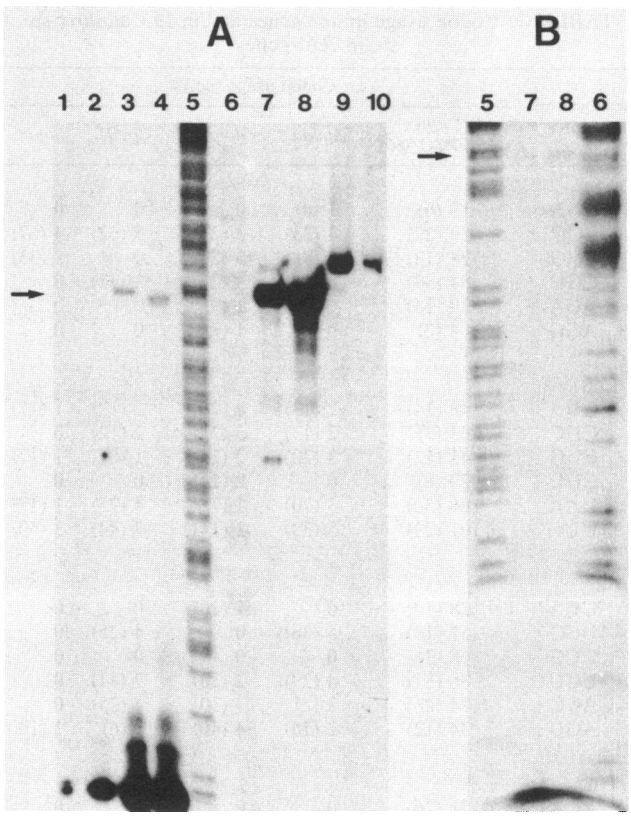

FIG. 6. Mapping of the 5′ extremities of the *apcA1B1C* (A) and *apcA2* (B) transcripts. Arrows indicate the positions of the S1-protected DNA fragments and of the reverse transcription products. Samples were run on 6% acrylamide sequencing gels. The *Dra*I-*Hin*cII fragments located either upstream from *apcA1* (303 bp [Fig. 2A]) or *apcA2* (112 bp [Fig. 2B]), 5′ labeled at the *Dra*I sites, were used to generate sequence ladders (A+G, lanes 5; C+T, lanes 6). (A) 303-bp fragment. Lanes: 10, before hybridization; 9, S1-untreated DNA-RNA hybridization mixture (RNAs were from cells grown in red light); lanes 7 and 8, S1-treated hybridization mixture with RNAs from cells grown in red light (lane 7) or in green light (lane 8). In the primer extension experiment, a *Dra*I-*Alu*I fragment (51 bp), 5′ labeled at the *Dra*I site, was used: primer alone (lane 1) or after extension with avian myeloblastosis virus reverse transcriptase following hybridization with *E. coli* tRNAs (lane 2), RNAs from cells grown in red light (lane 3), or RNAs from cells grown in green light (lane 4). (B) 112-bp fragment. Lanes: 7 and 8, S1-treated hybridization mixtures with RNAs from cells grown in red (lane 7) or green (lane 8) light.

the *apcA1B1C* operon, both S1 mapping and primer extension experiments gave the same results and allowed the identification of a unique and well-defined 5′ end for the transcripts (Fig. 6A). This 5′ end is located 193 nt upstream from the translation start site of *apcA1*. As defined by S1 mapping, transcription of the *apcA2* gene starts 98 nt upstream from the ATG (Fig. 6B). Unfortunately, the *apcA2*-specific primer could not be extended under our experimental conditions.

Since the 5′-flanking sequences of *apcA1* and *apcA2* share no homology, we designed more-specific probes, which only contained nonhomologous sequences. The *Dra*I-*Hin*cII 303-bp fragment, which ends 25 nt in front of the *apcA1* initiation codon, and the *Dra*I-*Hin*cII 112-bp fragment, which ends 21 nt in front of the *apcA2* gene, were used. The *apcA1*-specific probe revealed all but the 0.3-kb transcript. Under similar conditions, no transcripts were detected with the *apcA2* probe. However, after 8 days of exposure, a band about 0.65 kb could be seen on the fluoroautoradiogram, but bands corresponding to the 16S and 23S rRNA species were also detected at roughly the same intensity (data not shown).

From these results, we concluded that the three genes *apcA1*, *apcB1*, and *apcC* can be cotranscribed and that segmented transcripts originate from this gene cluster: the 1.7-kb species corresponds to the cotranscription of the three genes, the 1.4-kb species corresponds to that of *apcA1* and *apcB1* genes, and the 0.3-kb transcript corresponds to *apcC*. Finally, the occurrence of the mRNAs of about 5.6 kb indicates that the *apcA1B1C* gene cluster is part of a larger operon which includes gene(s) located upstream from the *apcA1* gene. As discussed below, the *apcE* gene (Fig. 2) was identified and partially sequenced. It encodes the L_{CM}^{92} and is 3.2 kb long (unpublished data). On the other hand, a monocistronic transcription unit most probably corresponds to the *apcA2* gene.

Nucleotide and amino acid sequence comparisons. The deduced polypeptide sequences for the four genes described here were compared with all of the corresponding known subunits from either cyanobacteria or eucaryotic organelles (Fig. 4). Regardless of the origin of these genes, a remarkably high degree of conservation was observed. The *apcA1* gene product of *Calothrix* strain 7601 was 73 to 83% identical to its homolog in the other organisms, whereas that of the *apcA2* gene was slightly less conserved (63 to 66%). Unexpectedly, less conservation exists between the gene products

inside the same strain: α^{AP1} and α^{AP2} were only 59% identical. The β^{AP} sequences were even more identical (82 to 93%) than were the α^{AP} sequences. Finally, the $L_C^{7.8}$ polypeptide encoded by *apcC* genes were 74 to 90% identical. It should be mentioned that most of the amino acid differences were conservative substitutions (K-R, S-T, D-E, or I-L-V-F-M) and that the α^{AP1}, β^{AP1}, and $L_C^{7.8}$ sequences were about 90% or more homologous to their counterparts in the other organisms, and the α^{AP2} sequences were about 75% homologous to those of the other α^{AP}s.

All of the α^{AP} subunits are likely to be 160 amino acids long, the size found by amino acid sequencing for the α^{AP}s of *Fischerella* sp. strain PCC 7603 (*Mastigocladus laminosus*), *Anabaena cylindrica*, and *Cyanidium caldarium*. The additional N-terminal methionine found when the α^{AP} sequences were deduced from nucleotide sequences is likely to be processed. In all cases, this methionine precedes a seryl residue which, according to Tsunasawa et al. (42), fits with the known specificity of methionine aminopeptidases. Similarly, the size of all the β^{AP} subunits is likely to be 161 residues. Processing indeed occurs in *Calothrix* strain 7601, as determined by direct amino acid sequence analysis of the polypeptide purified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (data not shown).

DISCUSSION

In all cyanobacteria examined so far, the two subunits (α and β) of the major phycobiliproteins (AP, PC, and PE) are encoded by adjacent and cotranscribed genes (3, 6, 11, 15, 25, 28, 29, 34). The genes specifying the linker polypeptides associated with either PC or AP are also found downstream from those coding for the PB with which they are associated. In some cases, these genes are cotranscribed (3, 25, 26; Tandeau de Marsac et al., in press) and give rise to segmented transcripts which all have the same 5′ end but different lengths. In good agreement with the stoichiometry observed within the PBsomes, the most-abundant transcripts are the shortest ones, which only encode the PB subunits. For PC and phycoerythrin, the gene encoding the β subunit is located 5′ to the gene which encodes the α subunit. In contrast, the opposite orientation (α followed by β) is found in the AP genes. The *Calothrix* strain 7601 *apcA1B1C* gene cluster has the same gene organization as that found with *Synechococcus* sp. strains PCC 6301 and PCC 7002, *Pseudanabaena* sp. strain PCC 7409, and *Nostoc* sp. strain PCC 8009 (25, 44; D. Bryant, *in* H. Scheer and S. Schneider, ed., *Photosynthetic Light-Harvesting Systems. Structure and Function*, in press). On the other hand, the *apcA2* gene resembles the *Calothrix* strain 7601 *apcD* gene in that no genes encoding PBsome components are found close to these genes.

S1 nuclease mapping indicates that transcription of the *apcA2* gene starts 98 bases upstream from the initiation codon. It is likely that the 3′ end of this transcript is located a few bases downstream from the highly stable stem and loop structure ($\Delta G = -29.7$ kcal/mol) which might be formed just following the end of the *apcA2* coding region. In agreement with the size suggested by the Northern blot experiments, a monocistronic unit of about 670 bases would thus correspond to the *apcA2* gene. Very few 3′ ends of cyanobacterial transcripts have been mapped so far. They all correspond to operons for which two or three segmented transcripts have been detected: *gvpA1A2C* (12), *cpcB1A1E* (29), and *cpcB2A2H2I2D2* (10, 26) from *Calothrix* strain 7601, as well as *cpcBACDE* from *Anabaena* sp. strain PCC 7120 (3). In every instance, 3′ ends were located a few bases downstream from stable stem and loop structures. Whether these structures act as a transcription terminator and/or as a barrier against 3′ exonuclease degradation remains to be determined. Multiple species also arise from transcription of the *apcEA1B1C* operon. The 1.4- and 1.7-kb transcripts have the same 5′ end but, in addition to the α and β subunits, the 1.7-kb species also encodes the $L_C^{7.8}$ polypeptide. The occurrence of a small transcript which only codes for the $L_C^{7.8}$ was not observed in *Synechococcus* sp. strain PCC 6301 (25). Whether it results from processing of the 1.7-kb species or from transcription initiated at an internal promoter located between *apcB* and *apcC* remains to be established. However, although the same transcripts have always been detected regardless of the light wavelength used to grow the cells, a difference in the relative abundance of the 1.7- and 0.3-kb transcripts was observed. The total amount of transcripts is about constant, but the 0.3- and 1.4-kb transcripts are three times more abundant and the 1.7-kb transcript is three times less abundant with regard to the total RNAs extracted from cells grown under red light than their counterparts in the RNAs extracted from cells grown in green light (Fig. 5). This observation leads us to favor the RNA-processing hypothesis. Interestingly, long transcripts were also detected as minor species. These RNAs must originate from sequences located upstream from *apcA1*, since probes located 3′ from the *apcC* gene failed to reveal any of the *apc* transcripts. Nucleotide sequence analysis of the *apcA1* upstream region revealed that an open reading frame exits (Fig. 1) which starts upstream from the *Eco*RI cloning site (data not shown). The 298 C-terminal residues of the corresponding polypeptide are more than 50% homologous at the amino acid level to the sequence of the open reading frame located upstream from *apcA* on the cyanelle genome of *Cyanophora paradoxa* (D. A. Bryant, personal communication). By analogy with the genetic organization of the *apc* genes recently reported by Zilinskas et al. (44), it is very likely that this gene product is the high-molecular-weight polypeptide (L_{CM}^{92}) located within the PBsome core and which, together with α^{APB}, acts as a terminal energy acceptor (20, 40). The corresponding gene has been designated *apcE* and is 3.2 kb long (data to be published elsewhere). The mRNA species of about 5.6 kb likely result from the transcription of the *apcEA1B1C* operon, including (upper band) or excluding (lower band) the *apcC* gene. Since transcripts which cover the entire operon are much less abundant than those which correspond to *apcA1B1C* (1.4 and 1.7 kb species), at least three possibilities can be envisaged: (i) RNA polymerase always starts transcription upstream from *apcE* but the 5′ region of such transcripts is far less stable than the 3′ region; (ii) only the long transcripts are made, but processing gives rise to mRNA species of highly different stability; and (iii) two promoters exist, one in front of the *apcEA1B1C* operon and a stronger one located between *apcE* and *apcA1*. At present, in *Calothrix* strain 7601, all of the genes encoding core components of PBsomes so far examined (*apcD*, *apcEA1B1C*, and *apcA2*) are transcribed independently of the light wavelength used to grow the cells. This observation is in agreement with the assumption that the synthesis of PBsome core components is not affected during complementary chromatic adaptation (40). In addition, no difference was observed in the 5′ end of the transcripts, suggesting constitutive expression from a unique promoter.

In *Calothrix* strain 7601, in contrast to the PC multigene family, which consists of three operons (each encoding both

α and β PC subunits), two *apcA* genes but only one *apcB* gene seem to exist. The α^{AP1} subunit is 73 to 83% identical to the amino acid sequences of the α^{AP} from other species and 72 to 77% identical at the nucleotide level (three of the six sequences have been deduced from nucleotide sequences [4, 25; D. Bryant, personal communication]). The α^{AP2} subunit is slightly less conserved (63 to 66% identity at the amino acid level and 64 to 66% identity at the nucleotide level). Surprisingly, α^{AP1} and α^{AP2} are only 59% identical at the amino acid level, but they are 67% identical at the nucleotide level; in contrast, the three α^{PC} and β^{PC} subunits of *Calothrix* strain 7601 are 84 to 87% and 75 to 78% identical at the amino acid level and 74 to 82% and 72 to 78% identical at the nucleotide level, respectively. The high level of nucleotide sequence identity observed between *apcA1* and *apcA2* covers the coding regions and extends to the putative ribosome-binding sites. Beyond that point, very little identity in both the 5′- and 3′-flanking regions is apparent (Fig. 3). Thus, within *Calothrix* strain 7601, the amino acid sequences of the α^{AP} subunits are less conserved than the nucleotide sequences, whereas values are about the same at both levels for the three α^{PC} and β^{PC} subunits. This result points out the occurrence of a selective pressure for nucleotide sequence conservation other than that imposed by functional constraints on the polypeptide chain. A biochemical characterization of the *apcA2* gene product and the elucidation of its role in the PBsome assembly would be of particular interest in finding a rationale for the maintenance of the *apcA2* gene copy despite its high degree of homology with the *apcA1* gene. Indeed, DNA sequence conservation following gene duplication increases recombination frequency between homologous sequences (2). The question thus arises: is the similarity between the *apcA1* and *apcA2* DNA sequences the result of evolutionary pressure (requirement for specific secondary structure of nucleic acids linked to either transcriptional or translational regulations or codon bias, for example), or is it the result of gene conversion? Gene conversion events have already been suspected to occur in *E. coli*, in which an almost-perfect stretch of homology has been observed between the *argF* and *argI* nucleotide sequences (43). Nevertheless, gene conversion in procaryotes is thus far poorly documented. Genetic exchanges between the multiple copies of the genome (two or more) present in cyanobacterial cells (35) or between "pseudoalleles" located on the same chromosome cannot be ruled out. On the *Calothrix* strain 7601 genome, three operons which encode PBsome components (*apcEA1B1C*, *cpcB2A2 H2I2D2*, and *cpcB1A1E*) are clustered (11), but the location of the other genes already characterized—namely, *cpcB3A3* (29), *cpeBA* (28), *apcD* (24), and *apcA2*—with regard to that cluster remains to be determined. Genetic exchanges could have occurred within the PC multigene family but would not have been detected because of the already very high degree of DNA sequence conservation. Gene conversion between the two *apcA* genes is less likely, since no significantly long stretches of perfect identity can be seen. The existence of possible secondary structures within the *apcA2* sequence, whose thermodynamic stability ranges from about −10 to −16 kcal/mol, might reflect constraints exerted for nucleotide sequence conservation.

Most of the genes which, in *Calothrix* strain 7601, encode the components identified in the core of the PBsomes have now been characterized: *apcD* (α^{APB}), *apcE* (L^{92}_{CM}), *apcA1* (α^{AP1}), *apcB1* (β^{AP1}), and *apcC* ($L^{7.8}_{C}$). The only missing gene is the one which codes for $\beta^{18.3}$, a β-type subunit that, like α^{APB} and L^{92}_{CM}, represents about 3% of the total PB subunit content of the core, i.e., 2 of 72 subunits. On the other hand, the polypeptide content of PBsome cores, as analyzed by two-dimensional gel electrophoresis, revealed the existence of multiple bands for some of the subunits (1, 19, 22). These results must now be reexamined in relation to the existence of multigene families in *Calothrix* strain 7601.

ACKNOWLEDGMENTS

We thank G. Cohen-Bazire, D. Mazel, T. Damerval, and G. Guglielmi for helpful discussions and support. We also thank D. A. Bryant for sharing his data with us prior to publication and M. Herdman for a critical reading of the manuscript.

This work was supported by the Centre National de la Recherche Scientifique (UA1129 and AI 990019) and by the Institut Pasteur.

ADDENDUM IN PROOF

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases under the accession numbers M20806 for the *apcA1B1C* operon and M20807 for the *apcA2* gene.

LITERATURE CITED

1. **Anderson, L. K., and F. A. Eiserling.** 1986. Asymmetrical core structure in phycobilisomes of the cyanobacterium *Synechocystis* 6701. J. Mol. Biol. **191**:441–451.
2. **Beacham, I. R.** 1987. Silent genes in prokaryotes. FEMS Microbiol. Rev. **46**:409–417.
3. **Belknap, W. R., and R. Haselkorn.** 1987. Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*. EMBO J. **6**:871–884.
4. **Bryant, D. A., R. de Lorimier, D. H. Lambert, J. M. Dubbs, V. L. Stirewalt, S. E. Stevens, Jr., R. D. Porter, J. Tam, and E. Jay.** 1985. Molecular cloning and nucleotide sequence of the α and β subunits of allophycocyanin from the cyanelle genome of *Cyanophora paradoxa*. Proc. Natl. Acad. Sci. USA **82**:3242–3246.
5. **Bryant, D. A., G. Guglielmi, N. Tandeau de Marsac, A. M. Castets, and G. Cohen-Bazire.** 1979. The structure of cyanobacterial phycobilisomes: a model. Arch. Microbiol. **123**:113–127.
6. **Capuano, V., D. Mazel, N. Tandeau de Marsac, and J. Houmard.** 1988. Complete nucleotide sequence of the red-light specific set of phycocyanin genes from the cyanobacterium *Calothrix* PCC 7601. Nucleic Acids Res. **16**:1626.
7. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4**:1539–1552.
8. **Cech, T. R., N. K. Tanner, I. Tinoco, Jr., B. R. Weir, M. Zucker, and P. S. Perlman.** 1983. Secondary structure of the *Tetrahymena* ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. Proc. Natl. Acad. Sci. USA **80**:3903–3907.
9. **Conley, P. B., P. G. Lemaux, and A. R. Grossman.** 1985. Cyanobacterial light-harvesting complex subunits encoded in two red light-induced transcripts. Science **230**:550–553.
10. **Conley, P. B., P. G. Lemaux, and A. Grossman.** 1988. Molecular characterization and evolution of sequences encoding light-harvesting components in the chromatically adapting cyanobacterium *Fremyella diplosiphon*. J. Mol. Biol. **199**:447–465.
11. **Conley, P. B., P. G. Lemaux, T. L. Lomax, and A. R. Grossman.** 1986. Genes encoding major light-harvesting polypeptides are clustered on the genome of the cyanobacterium *Fremyella diplosiphon*. Proc. Natl. Acad. Sci. USA **83**:3924–3928.
12. **Csiszàr, K., J. Houmard, T. Damerval, and N. Tandeau de Marsac.** 1988. Transcriptional analysis of the cyanobacterial *gvpABC* operon in differentiated cells: occurrence of an antisense RNA complementary to three overlapping transcripts. Gene **60**:29–37.
13. **Damerval, T., J. Houmard, G. Guglielmi, K. Csiszàr, and N. Tandeau de Marsac.** 1987. A developmentally regulated *gvpABC*

operon is involved in the formation of gas vesicles in the cyanobacterium *Calothrix* 7601. Gene **54**:83–92.
14. **De Lange, R. J., L. C. Williams, and A. N. Glazer.** 1981. The amino acid sequence of the β-subunit of allophycocyanin. J. Biol. Chem. **256**:9558–9566.
15. **De Lorimier, R., D. A. Bryant, R. D. Porter, W.-Y. Liu, E. Jay, and S. E. Stevens, Jr.** 1984. Genes for the α and β subunits of phycocyanin. Proc. Natl. Acad. Sci. USA **81**:7946–7950.
16. **Füglistaller, P., R. Rümbelli, F. Suter, and H. Zuber.** 1984. Minor polypeptides from the phycobilisome of the cyanobacterium *Mastigocladus laminosus*. Isolation, characterization and amino-acid sequences of a colourless 8.9-kDa polypeptide and of a 16.2-kDa phycobiliprotein. Hoppe-Seyler's Z. Physiol. Chem. **365**:1085–1096.
17. **Gantt, E.** 1981. Phycobilisomes. Annu. Rev. Plant Physiol. **32**: 327–347.
18. **Gantt, E., C. A. Lipschult, J. Grabowski, and B. K. Zimmerman.** 1979. Phycobilisomes from blue-green and red algae. Isolation criteria and dissociation characteristics. Plant Physiol. **63**:615–620.
19. **Gingrich, J. C., D. J. Lundell, and A. N. Glazer.** 1983. Core substructure in cyanobacterial phycobilisomes. J. Cell Biochem. **22**:1–14.
20. **Glazer, A. N.** 1985. Light harvesting by phycobilisomes. Annu. Rev. Biophys. Biophys. Chem. **14**:47–77.
21. **Glazer, A. N., D. J. Lundell, G. Yamanaka, and R. C. Williams.** 1983. The structure of a "simple" phycobilisome. Ann. Inst. Microbiol. (Paris) **134B**:159–180.
22. **Guglielmi, G., and G. Cohen-Bazire.** 1984. Etude taxonomique d'un genre de cyanobactéries oscillatoriacae: le genre *Pseudanabaena* Lauterborn. II. Analyse de la composition moléculaire et de la structure des phycobilisomes. Protistologica **XX**:393–413.
23. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
24. **Houmard, J., V. Capuano, T. Coursin, and N. Tandeau de Marsac.** 1988. Isolation and molecular characterization of the gene encoding allophycocyanin B, a terminal acceptor in cyanobacterial phycobilisomes. Mol. Microbiol. **2**:101–107.
25. **Houmard, J., D. Mazel, C. Moguet, D. A. Bryant, and N. Tandeau de Marsac.** 1986. Organization and nucleotide sequence of genes encoding core components of the phycobilisomes from *Synechococcus* 6301. Mol. Gen. Genet. **205**:404–410.
26. **Lomax, T. L., P. B. Conley, J. Schilling, and A. R. Grossman.** 1987. Isolation and characterization of light-regulated phycobilisome linker polypeptide genes and their transcription as a polycistronic mRNA. J. Bacteriol. **169**:2675–2684.
27. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Mazel, D., G. Guglielmi, J. Houmard, W. Sidler, D. A. Bryant, and N. Tandeau de Marsac.** 1986. Green light induces transcription of the phycoerythrin operon in the cyanobacterium *Calothrix* 7601. Nucleic Acids Res. **14**:8279–8290.
29. **Mazel, D., J. Houmard, and N. Tandeau de Marsac.** 1988. A multigene family in *Calothrix* sp. PCC 7601 encodes phycocyanin, the major component of the cyanobacterial light harvesting antenna. Mol. Gen. Genet. **211**:296–304.
30. **Minami, Y., F. Yamada, T. Hase, H. Matsubara, A. Murakami, Y. Fujita, T. Takao, and Y. Shimonishi.** 1985. Amino acid sequences of allophycocyanin α- and β-subunits isolated from *Anabaena cylindrica*. FEBS Lett. **191**:216–220.
31. **Mulligan, B., N. Schultes, L. Chen, and L. Bogorad.** 1984. Nucleotide sequence of a multiple-copy gene for the B protein of photosystem II of a cyanobacterium. Proc. Natl. Acad. Sci. USA **81**:2693–2697.
32. **Offner, G. D., and R. F. Troxler.** 1983. Primary structure of allophycocyanin from the unicellular rhodophyte, *Cyanidium caldarium*. The complete amino acid sequences of the α and β subunits. J. Biol. Chem. **258**:9931–9940.
33. **Parsot, C., and D. Mazel.** 1987. Cloning and nucleotide sequence of the *thrB* gene from the cyanobacterium *Calothrix* PCC 7601. Mol. Microbiol. **1**:45–52.
34. **Pilot, T. J., and J. L. Fox.** 1984. Cloning and sequencing of the genes encoding the α and β subunits of C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*. Proc. Natl. Acad. Sci. USA **81**:6983–6987.
35. **Roberts, T. M., L. C. Klotz, and A. R. Loeblich III.** 1977. Characterization of a blue-green algal genome. J. Mol. Biol. **110**: 341–361.
36. **Salser, W.** 1977. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **77**:985–1010.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Sidler, W., J. Gysi, E. Isker, and H. Zuber.** 1981. The complete amino acid sequence of both subunits of allophycocyanin, a light-harvesting protein-pigment complex from the cyanobacterium *Mastigocladus laminosus*. Hoppe-Seyler's Z. Physiol. Chem. **362**:611–628.
39. **Stanier, R. Y., and G. Cohen-Bazire.** 1977. Phototrophic prokaryotes: the cyanobacteria. Annu. Rev. Microbiol. **31**:225–274.
40. **Tandeau de Marsac, N.** 1983. Phycobilisomes and complementary chromatic adaptation in cyanobacteria. Bull. Inst. Pasteur **81**:201–254.
41. **Tandeau de Marsac, N., D. Mazel, D. A. Bryant, and J. Houmard.** 1985. Molecular cloning and nucleotide sequence of a developmentally regulated gene from the cyanobacterium *Calothrix* PCC 7601: a gas vesicle protein gene. Nucleic Acids Res. **13**:7223–7236.
42. **Tsunasawa, S., J. W. Stewart, and F. Sherman.** 1985. Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. J. Biol. Chem. **260**:5382–5391.
43. **van Vliet, F., R. Cunin, A. Jacobs, J. Piette, D. Gigot, M. Lauwereys, A. Piérard, and N. Glansdorff.** 1984. Evolutionary divergence of genes for ornithine and aspartate carbamoyltransferases—complete sequence and mode of regulation of the *Escherichia coli argF* gene: comparison of *argF* with *argI* and *pyrB*. Nucleic Acids Res. **12**:6277–6289.
44. **Zilinskas, B. A., K. H. Chen, and D. A. Howell.** 1987. Cloning genes for phycobilisome core components. Plant Physiol. Suppl. **83**:60.