The presence of multidomain linkers determines the bundle-shape structure of the phycobilisome of the cyanobacterium Gloeobacter violaceus PCC 7421

ARTICLE *in* PHOTOSYNTHESIS RESEARCH · AUGUST 2007

Impact Factor: 3.5 · DOI: 10.1007/s11120-007-9133-9 · Source: PubMed

Photosynth Res (2007) 93:27–43
DOI 10.1007/s11120-007-9133-9

ORIGINAL PAPER

The presence of multidomain linkers determines the bundle-shape structure of the phycobilisome of the cyanobacterium *Gloeobacter violaceus* PCC 7421

David W. Krogmann · Bertha Pérez-Gómez · Emma Berta Gutiérrez-Cirlos · Alicia Chagolla-López · Luis González de la Vara · Carlos Gómez-Lojero

Received: 25 July 2006 / Accepted: 7 January 2007 / Published online: 20 February 2007

Abstract The complete genome sequence of *Gloeobacter violaceus* [Nakamura et al. (2003a, b) DNA Res 10:37–45, 181–201] allows us to understand better the structure of the phycobilisomes (PBS) of this cyanobacterium. Genomic analysis revealed peculiarities in these PBS: the presence of genes for two multidomain linker proteins, a core membrane linker with four repetitive sequences (REP domains), the absence of rod core linkers, two sets of phycocyanin (PC) α and β subunits, two copies of a rod PC associated linker (CpcC), and two rod cap associated linkers (CpcD). Also, there is one ferredoxin–$NADP^+$ oxidoreductase with only two domains. The PBS proteins were investigated by gel electrophoresis, amino acid sequencing and peptide mass fingerprinting (PMF). The two unique multidomain linkers contain three REP domains with high similarity and these were found to be in tandem and were separated by dissimilar Arms. One of these, with a mass of 81 kDa, is found in heavy PBS fragments rich in PC. We propose that it links six PC hexamers in two parallel rows in the rods. The other unique linker has a mass of 91 kDa and is easily released from the heavy fragments of PBS. We propose that this links the rods to the core. The presence of these multidomain linkers could explain the bundle shaped rods of the PBS. The presence of 4 REP domains in the core membrane linker protein (129 kDa) was established by PMF. This core linker may hold together 16 AP trimers of the pentacylindrical core, or alternatively, a tetracylindrical core of the PBS of *G. violaceus*.

Keywords Cyanobacterium · *Gloeobacter violaceus* · Phycobilisome · Multidomain linkers

Abbreviations

AP	Allophycocyanin
apcAB	Gene encoding the α and β subunits of allophycocyanin
apcC	Gene encoding the 8–kDa core linker polypeptide
ApcE	The 129-kDa core membrane linker phycobiliprotein
A.U.	Absorbance units
cpcBA	Genes encoding the α and β subunits of phycocyanin
cpcC1	Gene encoding the PC-rod linker polypeptide
cpcD	Gene encoding the 9-kDa rod linker polypeptide
cpcG	Gene encoding the rod-core linker polypeptide
DCPIP	Dichlorophenol-indophenol

D. W. Krogmann
Department of Biochemistry, Purdue University, West Lafayette, IN 47907-1157, USA

B. Pérez-Gómez · C. Gómez-Lojero (✉)
Departmento de Bioquímica, Centro de Investigación y Estudios Avanzados-IPN, Apartado Postal 14-740, 07000 Mexico, D.F., Mexico
e-mail: cgomez@cinvestav.mx

E. B. Gutiérrez-Cirlos
Unidad de Biomedicina FES-Iztacala UNAM, Av. De los Barrios 1, Los Reyes Iztacala, Tlalnepantla, Edo de México 54090, México

A. Chagolla-López · L. González de la Vara
Cinvestav-Unidad Irapuato, Apartado Postal 629, 36503, Guanajuato, Mexico

DMF	Dimethyl-formamide
EDTA	Ethylenediamine tetraacetic acid
FNR	Ferredoxin-NADP$^+$ oxidoreductase
FNR-2D	FNR containing only the two catalytic domains
FNR-3D	FNR containing three domains (the CpcD-like domain and two catalytic domains)
F.U.	Fluorescence units
L_C^8	Core linker polypeptide of 8 kDa
L_{CM}^{129}	L_R^9, the 9-kDa, phycocyanin-associated rod linker polypeptide
L_R^{33}	CpcC, the PC-rod linker polypeptide
CpcG	The rod core linker polypeptide
PAGE	Polyacrylamide gel electrophoresis
PBP(s)	Phycobiliprotein(s)
PBS	Phycobilisome(s)
PC	Phycocyanin
PE	Phycoerythrin
petH	Gene encoding FNR
PMF	Protein mass fingerprinting
PMSF	Phenyl-methylsulfonyl fluoride
REP	Repetitive sequence
SDS	Sodium dodecylsulfate
WCE	Whole-cell extract

Introduction

Phycobilisomes (PBSs) are the major light-harvesting systems in cyanobacteria and red algae. This type of photosynthetic antenna consists of a macro complex composed of light-absorbing proteins named phycobiliproteins (PBPs). The PBPs are composed of α and β subunits and are a brilliantly colored group of disc-shaped proteins bearing covalently attached phycobilins. In addition, linker proteins hold together the PBPs. Both PBPs and linker proteins are organized as a core connected with photosystems and surrounded by several peripheral rods (Sidler 1994; MacColl 1998; Tandeu de Marsac 2003; Grossman 2003; Adir 2005; Liu et al. 2005). *G. violaceus* is a unicellular cyanobacterium, which lacks thylakoids. It has a lower relative ratio of chlorophyll *a* with phycobiliproteins than is found in most typical cyanobacteria (Rippka et al. 1974). Underlying the membrane is an electron-dense layer 80-nm thick, which seems to be associated with the PBSs. *G. violaceus* PBSs differ in their structural organization from those of other cyanobacteria and red algae (Guglielmi et al. 1981): the core is described as amorphous and the rods containing phycocyanin and phycoerythrin are grouped as a bundle. These bundles are organized as two rows of three parallel rod-shaped elements. The rods are 10–12 nm wide and their length varies from 50 to 70 nm. These rods appear to be formed by the coaxial stacking of 7–10 discs, 6-nm thick and 12 nm in diameter (Guglielmi et al. 1981). In addition, *G. violaceus* PBSs are considerably longer than the peripheral rods of hemidiscoidal PBS. Therefore, the PBSs of *G. violaceus* have a complex composition since they also possess three types of constitutive PBPs: allophycocyanin (AP), phycocyanin (PC) and phycoerythrin (PE) (Bryant et al. 1981). *G. violaceus* does not adapt chromatically, and therefore the proportions of the three phycobiliproteins remain unchanged after growing in white, red or green light (Rippka et al. 1974). The genes encoding the phycobiliprotein subunits and the linker proteins of the PBS of *G. violaceus* have been recently identified in the complete genome of this organism (Nakamura et al. 2003a). The genome contains at least 33 open reading frames (ORFs) encoding polypeptides with similarity to proteins involved in PBS assembly, phycobilin biosynthesis or ligation, and PBS regulation (Nakamura et al. 2003b). The genes may be grouped in five main clusters as seen in Fig. 1. Several unusual features are found in the PBS-related genes. The AP gene family contains a cluster *apcEABC*. In addition, two genes are found isolated in the genome, a second gene for *apcA* and an *apcD* gene (encoding α^{APB}). The gene encoding $\beta^{AP\text{-}18}$ is absent from the genome. The *apcE* gene, encoding the PBS core-membrane linker polypetide (L_{CM}), contains four repetitive sequences (REP domains, Capuano et al. 1991) similar to those found in *Anabaena* sp. PCC 7120 (Kaneko et al. 2001), *Mastigocladus laminosus* (Sidler 1994) and *Thermosynechococcus* (Nakamura et al. 2002). A single gene cluster, designated *cpeBA* encodes both PE subunits. There are three genes encoding proteins with similarity to the PE linkers: *cpeC*, *cpeD* and *cpeE* making a cluster *cpeCDE* that is separated from *cpeBA* as in *Fremyella diplosiphon* (Cobley et al. 2002). Two distinct gene sets encode PC subunits; these have been designated *cpcB1A1* and *cpcB2A2*. Furthermore, two genes encoding the linker polypeptide associated with PC have been designated *cpcC1* and *cpcC2* (Gutiérrez-Cirlos et al. 2006). These two genes are also isolated in the genome. Two genes encoding the rod-terminating linkers have been designated *cpcD1* and *cpcD2*, and are located in the same cluster as the PE associated linkers. The *cpcG* genes, encoding for the core rod

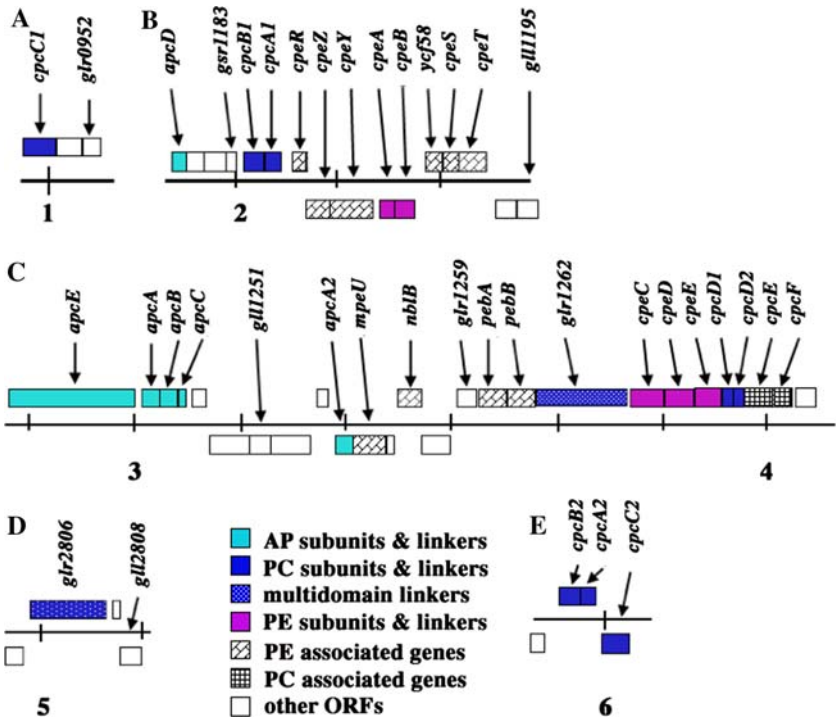

Fig. 1 Complete map of phycobilisome genes from *G. violaceus* PCC 7421. The numbers below the bars (1–6) give the positions of the five clusters in the genome. Number 1 is located at nt 1,000,000, number 2 is at nt 1,263,000, number 3 is at nt 1,326,000, number 4 is at nt 1,344,000, number 5 is at nt 2,985,000 and number 6 is at nt 3,426,000. The *cpcB* and *cpcA* genes are frequently followed by genes encoding PC-linker polypeptides in other cyanobacteria (Tandeau de Marsac et al. 1988). The genes encoding PC-linker polypeptides (dark blue) are clearly scattered in different regions. Also, the novel genes for multidomain linkers of this cyanobacterium (*glr1262* and *glr2806*) are distant from one another

linker proteins (Bryant et al. 1991) are absent from the genome (Nakamura et al. 2003a). Finally, the *petH* gene, which encodes ferredoxin–NADP oxidoreductase (FNR) predicts a protein of only two domains (296 amino acids), and lacks a CpcD-like domain at its N-terminus. FNR is not expected to serve as a component of the PBS (Schluchter and Bryant 1992; van Thor et al. 1999; Gómez-Lojero et al. 2003).

The linker proteins determine the positions of the phycobiliproteins within the phycobilisome structure (Bryant et al. 1991; Glauser et al. 1992). Also, the linker proteins interact either directly or indirectly with the chromophores causing changes in their environment. These changes can modulate the spectral properties of different phycobilisome subassemblies (Gottschalk et al. 1991, 1993; Reuter et al. 1999). The rod linker proteins of *G. violaceus*, CpcC1, CpcC2, CpeC, CpeD and CpeE, have been identified by N-terminal sequencing (Gutiérrez-Cirlos et al. 2006). Two of them are rod phycocyanin associated linker proteins and three are rod phycoerythrin associated linker proteins. The genome sequence of *G. violaceus* predicts the presence of unique components in the PBS of this cyanobacterium. In this work, we have been able to confirm the presence of linkers in intact and dissociated PBSs from *G. violaceus.* Using a biochemical approach, we characterized some of the linker properties, which allow us to suggest their localization within the PBS.

Materials and methods

The sequences of the genome of *G. violaceus* were retrieved from GenBank, accession number BA000045. The online programs for Blastp analyses (Altschul et al. 1997) were obtained using NCBI and ExPASy Web sites. Alignments of linker polypeptides were made using the ClustalX (Thompson et al. 1997). The percentages of identity and similarity between two sequences were obtained after Blastp analyses. Sequences of PBS proteins from other cyanobacteria were retrieved from Entrez (GenBank NCBI).

Membrane extraction and centrifugation in sucrose gradients

G. violaceus (Rippka et al. 1979) was supplied by the Pasteur culture collection. Cultures were grown in 50 ml of BG 11 medium supplemented with 5 mM Mops (pH 7.8). A light intensity of 5 μmol m^{-2} s^{-1} was employed. The cultures were grown for 18–25 days. The cells were washed once with 0.65 M Na^+/K^+ phosphate buffer (pH 8.0) and suspended in the same buffer. The cells were disrupted once or twice in a French pressure cell (Thermo Spectronic) operated at 20,000 psi. The broken cells were centrifuged for 10 min at 27,000 × *g*. The supernatant of whole cell extract (WCE) was treated with Triton X-100 (1–1.2% v/v final concentration) for 30 min and centri-

fuged again for 30 min at 27,000 × *g*. Part of the supernatant was carefully layered on the sucrose gradient. In an alternative experiment, 240 mg per ml ammonium sulfate was added to the colored supernatant and the mixture was centrifuged at 27,000 × *g* for 30 min. The precipitate was suspended with 0.65 M Na^+/K^+ phosphate buffer (pH 8.0). About 2 ml samples were layered onto sucrose gradient tubes. Alternatively, when indicated, the broken cells were centrifuged at 150,000 × *g* for 1 h to sediment the membranes. The membranes were suspended in the same buffer as above, solubilized with Triton X-100 (1%) and layered onto the sucrose gradient. The sucrose step gradients consisted of 4, 5, 4, and 3 ml of 0.5, 0.75, 1.0 and 2.0 M. sucrose solutions in 0.75 M Na^+/K^+ phosphate buffer pH 8.0. The gradient was centrifuged in a Beckman Ti 60 rotor at 65,000 × *g* for 12 h at 20°C or at 225,000 × *g* for 5 h at 20°C. The gradients were collected from bottom to top in 1 ml aliquots. The sucrose concentration of each fraction was determined with an Abbe refractometer. Alternatively, the color fractions were pooled and each band pool was precipitated with TCA (10% final concentration) for subsequent SDS—or lithium dodecyl sulfate (LDS)—PAGE analysis. Absorption spectra of all fractions were obtained by scanning from 340 nm to 700 nm and the absorbances at 502, 620 and 656 nm were used to calculate the amounts of PE, PC and AP. Fluorescence spectra were recorded with a Hitachi F4500 fluorescence spectrophotometer.

SDS and LiDS-PAGE

SDS-PAGE analyses were performed on 10% or 14% polyacrylamide slab gels (Schägger and von Jagow 1987). LiDS-PAGE was performed on 10% polyacrylamide slab gels (Delepelaire and Chua 1979). Electrophoresis was run at 4°C and at constant voltage for 12 h. For fluorescence visualization under UV light, gels were pre-incubated for 15 min in 20 mM zinc acetate (Raps 1990). The proteins were identified by PMF. The bands corresponding to the proteins stained with Coomassie blue were excised from the gel and sent to the Protein Identification Unit of Cinvestav at Irapuato, México. After in-gel trypsin digestion (using sequence grade trypsin from PROMEGA), proteins were analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI ToF MS). The experimental mass values were compared with those derived from available databases and each subunit was then identified in the same instrument (Ettan MALDI-ToF) or by using the MASCOT program.

Densitometric scans were performed with a Pharmacia LKB UltroScan XL and analyzed with the program GSXL version 2.1. The relative molar amounts of protein per band were obtained as described (Gómez-Lojero et al. 1997). The scaling factors used were 1.29 for L_{CM}^{120}; 0.91 for L_{RC}^{91} (Glr1262) 0.81 for L_R^{81}; 0.33 for L_R^{33}, 0.30 for L_R^{30}, 0.35 for L_R^{35}.

N-terminal sequence analysis

The proteins were prepared for N-terminal sequence analyses as follows. The SDS-PAGE was run with 150 μg of protein in each well of a 1.5-mm thick slab of 10%-polyacrylamide gel for high molecular weight proteins and 14% polyacrylamide gel for low molecular weight proteins. The proteins were then transferred to a Trans Blot (Bio Rad PVDF) membrane using the Trans Blot SD (Bio Rad) semi-dry system. The PVDF membrane was stained with Coomassie brilliant blue R-250 and the bands were cut and sent to the Protein Analysis Laboratory at Purdue University for N-terminal sequencing.

Molecular mass estimation of the PBS-substructures

The estimation of molecular mass was performed on sucrose density gradients by an established method (Martin and Ames 1961). Linear gradients, (16 ml), of 0.5–2 M sucrose in 0.75 M Na^+/ K^+ phosphate, pH 8.0 were prepared and 2 ml of protein solution was layered on each gradient. Centrifugation was performed in a Beckman 60 Ti rotor at 225,000 × *g* for 5 h at 20°C. About 1 ml fractions were collected with a peristaltic pump from the bottom to the top of the tube. Phycobiliprotein distribution in the gradient was determined by the absorbance at the peak of each protein. The cells were disrupted only once in the French press, treated with Triton X-100 (1%) for 30 min and then centrifuged for 20 min at 27,000 × *g*. The supernatant was precipitated with ammonium sulfate (40% saturation). The suspended precipitate (2 ml) was layered on the sucrose gradient.

Absorption and fluorescence spectroscopy

Absorption spectra were recorded with an Aminco DW2 UV–visible spectrophotometer with the OLIS DW2 conversion and OLIS software or with a Hitachi U3310 spectrophotometer. The concentrations of phycobiliproteins were estimated using the extinction coefficients for the protomers of *G. violaceus* $\varepsilon^{PE}{}_{502}$ 245, $\varepsilon^{PE}{}_{566}$ 456, $\varepsilon^{PC}{}_{566}$ 141, $\varepsilon^{PC}{}_{620}$ 365, $\varepsilon^{PC}{}_{654}$ 51.5

(Bryant et al 1981), $\varepsilon^{AP}{}_{566}$ 42, $\varepsilon^{AP}{}_{620}$ 172, $\varepsilon^{AP}{}_{654}$ 313 (Füglistaller et al. 1987; Ducret et al. 1998), and are expressed in $mM^{-1}\ cm^{-1}$. Pigment concentrations were calculated with the following set of simultaneous equations:

$$[AP] = (A_{654} - 0.141A_{620}) \div 289;$$
$$[PC] = (A_{620} - 0.55A_{654}) \div 337;\quad [PE] = A_{502} \div 245$$

Fluorescence spectra were recorded with a Hitachi F–4500 fluorescence spectrophotometer.

Results

Identification of PBS subunits

The distribution of the *G. violaceus* PBPs after sucrose gradient centrifugation was characterized spectroscopically. Three main fractions of PBPs (3, 6, and 11) absorbing at 566 nm were identified (Fig. 2A). The most rapidly sedimenting fraction contained the PBSs. More than 55% of the pigment was found as PBS (fractions 2–4). Figure 2B shows the spectrum of fraction 3 (PBS), from which 1 to 2.15 to 1.7 molar ratios of AP absorbance at 654 nm, PC absorbance at 620 nm and PE absorbance at 502 nm were calculated. The ratios of the PBPs are close to those found in the whole cell extract (1 : 2.2 : 1.66). The middle peak contained a substructure enriched in PE and PC. The spectrum of fraction 6 is shown in Fig. 2C, and exhibits ratios of AP to PC to PE of 1 to 2.3 to 2.6. The slowly sedimenting peak absorbing at 566 nm is congruent with the peak absorbing at 620 nm, with a ratio of PC to PE of 4 and was obtained from the spectrum of fraction 11 of Fig. 2C. The sucrose gradient shown in Fig. 2A was linear from 15% to 42%, and thus allowed us to estimate the relative sizes of some of the resulting substructures. The ratio of the distances traveled from the top to fraction 11 and to fraction 13 is 1.54, which corresponds to a molecular weight ratio of 1.9, $R = (MW_1/MW_2)^{2/3}$ (Martin and Ames 1961). We assigned the molecular mass of 226,794 Da to fraction 11 (calculated for PE and PC hexamers) and thus a calculated mass of 119 kDa was obtained for fraction 13, which contains trimers of AP. The spectrum of fraction 5 in Fig. 2C revealed two main substructures: one rich in AP and the other in PC. A mass of 1,500,000 Da was assigned to fraction 5, whose molar ratios of biliproteins, were 1 : 1.15 : 1 for AP to PC to PE. The estimated molecular mass for an intact pentacylindrical core is 2,000,000 Da (Ducret et al. 1998), and for six PC hexamers is 1,451,000 Da. We concluded that fraction 5 contained a partially dissociated pentacylindrical core and a substructure formed by six PC hexamers.

In order to characterize the proteins of the *G. violaceus* PBS and of its corresponding substructures, the colored bands from the gradients were pooled from the ultracentrifuge tubes and subjected to SDS-PAGE analyses. Figure 3 shows two electrophoretograms of the colored bands obtained from the sucrose gradient shown in Fig. 2 Panel D. The electrophoretogram on the left shows a 10% acrylamide gel, used to identify the high molecular weight components, while on the right electrophoretogram there is a 14% acrylamide gel to identify the low molecular weight components. Lanes I to III in both gels display the proteins in the colored bands obtained from the sucrose gradient shown in Fig. 2D. The apparent molecular masses of the PBS polypeptides determined by SDS-PAGE, the molecular masses predicted from amino acid sequences, and the method used for identifying each protein are presented in Table 1. Fraction I (left electrophoretogram on Fig. 3) contained three linker proteins of 129, 91 and 81 kDa, which correspond to ApcE ($L_{CM}{}^{129}$) and to the protein products coded by *glr1262* and *glr2806* genes, respectively. Fraction II exhibited a higher proportion of the 129 and 81 kDa polypeptides and less of the 91 kDa polypeptide as compared with Fraction I (the quantitative aspects of these electrophoretograms are presented in Table 2). The polypeptides from Gll0415 and Glr0638 are conspicuous in lane I, vary from sample to sample and were identified as porins that are the main proteins of the external membrane. The groups of polypeptides of 35, 33 and 30 kDa have been previously identified as CpeC, CpcC1, CpcC2, CpeD and CpeE (Gutiérrez-Cirlos et al. 2006). Lane III of this gel was used to correlate the quantities of the middle size linkers with their associated PBPs. The 14% acrylamide gel (at the right of the Figure) was used to identify the low molecular weight components. In lane I, a faint band of 8 kDa could be seen at the bottom of the gel. In lane II, two faint bands of 12 and 11 kDa were detected. The intense band of 8 kDa corresponds to ApcC as shown by PMF (Table 1). Fraction III contained more intense subunits of 12 and 11 kDa, which were identified as CpcD1 and CpcD2 by N-terminal sequence analysis (Table 1). Together the results of both 10 and 14% gels led us to conclude that the PBS of *G. violaceus* contains 11 linker proteins (Table 1). Changes in concentration of other proteins are also shown in Table 2 and are correlated with the percentage of PBPs obtained from the spectra of the bulk fractions shown in Fig. 2D. Data

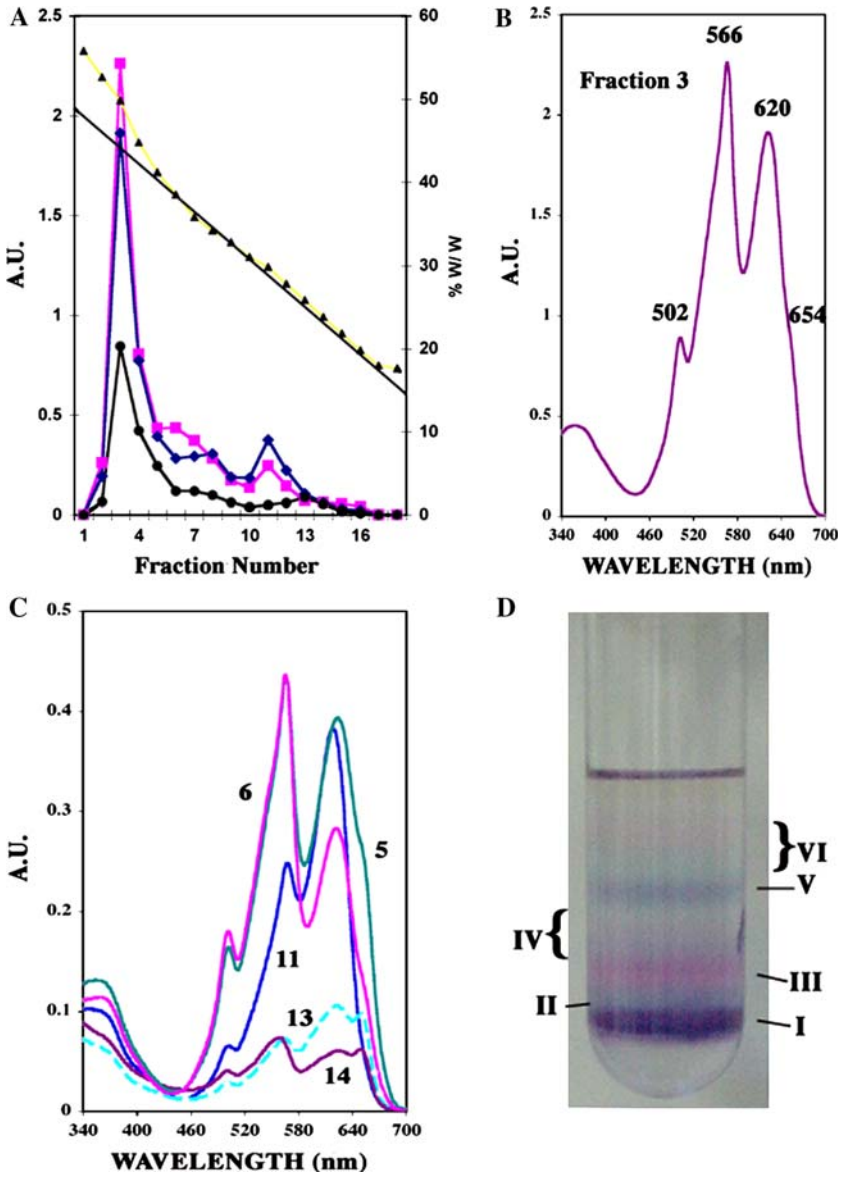

Fig. 2 Characterization of the PBS and PBS substructures. The WCE was obtained after one pass through the French press, solubilized by Triton X-100 and precipitated by ammonium sulfate. The sucrose gradient (described in Materials and methods) was loaded with the suspended ammonium sulfate precipitate and centrifuged for 5 h at 225,000 x *g*. Panel **A** shows the pattern of fractions (bottom to top) obtained from the sucrose gradient. Tubes containing 1 ml fractions were collected. The absorbance of each fraction at 566 nm for PE (*fuchsia squares*), 620 nm for PC (*blue diamonds*) and 654 nm for AP (*black circles*) are indicated. The fractions were assayed for sucrose concentration (*triangles*) to estimate the molecular weights of the PBS substructures. Panel **B**: Absorption spectrum of the PBS (fraction 3 of Panel **A**). Panel **C**: Absorption spectra of fractions 5, 6, 11, 13, and 14 after the sucrose gradient of Panel **A**. Panel **D**: The sucrose gradient showing the bulk colored fractions pooled for SDS-PAGE analyses

in Table 2 allow us to establish a correlation between the linkers in the different fractions of the sucrose gradient and the amount of each of the PBPs. The high molecular weight linkers L_{CM}^{129}, L_{RC}^{91} and L_R^{81} are present in the heavy fractions. The middle-size linkers L_R^{35}, L_R^{33} and L_R^{30} correlate with PC and PE. As expected L_R^{30} correlates with PE and L_R^{33} with PC. In contrast, L_R^{35} correlates in a more complex manner with two components, one that associates PC with PE, and another that interacts with two PC hexamers. The L_C^8 correlates with AP. The L_R^{11} and L_R^{12}, which are the terminating linkers, are detected in the fraction rich in PE that is the distal part of the rods (Lanes II and III in Fig. 3 and Table 2).

Localization of the linker proteins

G. violaceus membranes, free from soluble dissociated PBS substructures, were solubilized with Triton X-100 and applied to a sucrose gradient. A different pattern of PBPs was obtained in this linear gradient. In particular, the blue band between the fuchsia and the violet bands was clearly observed. We characterized both the violet and the blue bands. Figure 4A shows the absorption and the fluorescence excitation spectra of the violet band. The absorption spectrum shows PBS with almost the same absorption by phycocyanin and phycoerythrin. The minor peak at 502 nm corresponds to the phycourobilin, one of the chromophores of PE. The near equality of the maxima in the spectrum

Table 1 Phycobilisome linker proteins and co-migrating porins of *G. violaceus* as obtained from genomic analysis and classified according to OLN (order locus name)

ORF No	Gene	Predicted Molecular mass (Da)	SDS-PAGE M.w. (kDa)	N-terminal sequence	PMF No	PMF coverage	Proposed function
gll415		66,045	63		12	24%	PORIN
glr638		62,027	59		12	31%	PORIN
glr0950	*cpcC1*	31,043	35	Gutierrez-Cirlos et al. 2006			L_R^{31}
glr1245	*apcE*	129,836	120*		24	24%	L_{CM}^{129}
glr1248	*apcC*	7,754	8		3	52%	$L_C^{7.7}$
glr1262		91,882	87	VLDDQSKAY			L_{RC}^{91}
glr1263	*cpeC*	31,825	35	Gutierrez-Cirlos et al. 2006			$L_R^{31.8}$
glr1264	*cpeD*	28,400	30	Gutierrez-Cirlos et al. 2006			$L_R^{28.4}$
glr1265	*cpeE*	28,354	30	Gutierrez-Cirlos et al. 2006			$L_R^{28.3}$
glr1266	*cpcD1*	8,169	12	SGXVXTGVAX			$L_R^{8.1}$
glr1267	*cpcD2*	7,769	11	MYXNVTAXXX			$L_R^{7.7}$
glr2806		81,442	81	SATTYDWRKVI			L_R^{81}
gll3219	*cpcC2*	30,878	33	Gutierrez-Cirlos et al. 2006			$L_R^{30.8}$

The characterizing abbreviations are given at the top of each column. PMF coverage indicates the percentage of the amino acid residues identified in fragments by PMF/total amino acids residues. Letter X in the N-terminal sequences of $L_R^{8.1}$ (Glr1266) and $L_R^{7.7}$ (Glr1267) indicates that the predominant amino acid found is not the one expected from the genome sequence. However, the expected amino acid was present in a lower concentration. Despite the ambiguity in these positions, the remainder of the sequences provided the unambiguous identification of CpcD1 and CpcD2 in the *G. violaceus* genome

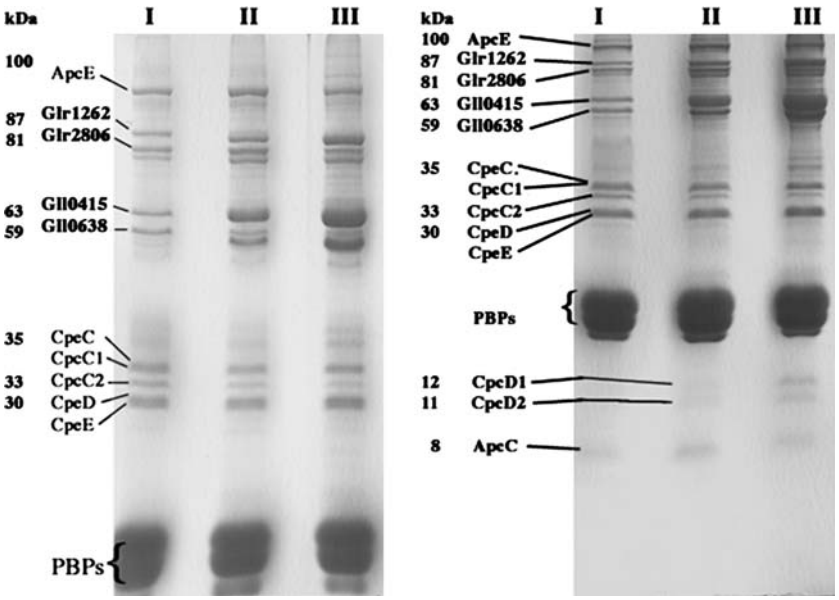

Fig. 3 LiDS-PAGE analyses of the pooled bands (I–III) of the sucrose gradient shown in Fig. 2, Panel D. Bands IV–VI were also run but the electrophoretograms are not shown. The left Panel shows a 10% acrylamide gel. The electrophoresis was stopped when the PBPs arrived at the bottom edge of the gel for the proper separation of the high molecular weight components. The right Panel shows a 14% acrylamide gel for detection of the low molecular weight components. This electrophoresis was stopped when the bromophenol blue arrived at the bottom edge of the gel. Each of these gels was prepared with duplicates of each of the pooled fractions. One half of the gel was blotted onto a PVDF membrane to yield bands for N-terminal sequencing. The other half was used for Coomassie blue staining and densitometric scanning, to obtain molecular masses, the relative amounts of the proteins and, finally, for PMF. The proteins identified by N-terminal sequencing and by PMF are shown at the left of each electrophoretogram. The gels were scanned and the estimated molecular masses (molecular weight markers were included in each gel, but not shown) are reported in Table 1, the calculated relative amounts of proteins are presented in Table 2

Table 2 Relative amounts of the components of PBS and PBS substructures of *G. violaceus* isolated from the sucrose density gradients

Subunit	I (10)	I (14)	II (10)	II (14)	III (10)	III (14)
L_{CM}^{129}	4	4	6	7	2	
L_{RC}^{91}	4	3	2		0	
L_{R}^{81}	5	4	9	7	5	
L_{R}^{35}	16	15	14	12	14	16
L_{R}^{33}	8.2	8	5	4	3	4
L_{R}^{30}	25	21	25	26	30	34
PBPs	100	100	100	100	100	100
AP	21	21	32	32	18	18
PC	42	42	38	38	40	40
PE	37	37	30	30	42	42
L_{R}^{12}				6		9
L_{R}^{11}				5		7
L_{C}^{8}		3		40		10

In order to obtain quantitative values of intensities of the bands from the proteins stained with Coomassie blue, the gels shown in Fig. 3 were scanned by laser densitometry. The area under the curve (A.U. × mm) obtained from the densitometric scan was multiplied by the factor related to its molecular weight and described in Materials and methods. The above result for the PBPs was divided by 12 (the number of subunits per PBP hexamer). By using this number all the values of the linkers were normalized, taking the value for the PBPs as 100. The percentage of each PBP (AP, PC and PE) was obtained by calculating the concentration from absorbance values of the spectra of each of the fractions employed (I–III) from the sucrose gradient. Each of the linker components is listed with its proper linker name and its molecular weight in the superscript. The PBP components are listed as well. Numbers in parentheses indicate the percentage of the acrylamide gel used for the determination

indicates the loss of some PE during the first stage of the preparation. The fluorescence excitation spectrum reveals energy transfer from PE and PC to AP since the emission wavelength used is 687 nm, and is indicative of energy transfer to AP, which is the final energy acceptor in the PBS. Inserted in Panel A is a LiDS-PAGE that shows the components of this PBS. Also, the ratios of high molecular weight components of the PBS are inserted beneath the spectra. Figure 4B contains information about the blue band above the PBS. The absorption spectrum shows that PC and AP are the main components of this fraction. The fluorescence excitation spectrum shows less energy transfer with a maximum at 665 nm corresponding to the excitation of the last AP component of the PBS. From the inserted LiDS-PAGE in Panel B, we obtained data about the ratios of the high molecular weight linkers, shown beneath the spectra. The concentration of L_{RC}^{91} was particularly low in this fraction or null in other experiments. We concluded that this fraction is formed by the core of the PBS and probably by six PC hexamers maintained together as a substructure by L_{RC}^{81}. The above interpretation is in agreement with the molecular weights estimated from Fig. 2.

Figure 5A shows gels of the high molecular mass linker polypeptides of *G. violaceus* from the violet (I), blue (II) and fuchsia (III) bands obtained from the sucrose gradient like those seen in Fig. 2D and compares them with L_{CM}^{99} of *Synechococcus* sp. PCC 7002 (marked as 7002). Panel B shows the same region after treatment with zinc acetate and UV transillumination. Using the known molecular mass of *Synechococcus* L_{CM}^{99}, as a marker, an apparent molecular mass of 120 kDa for the *G. violaceus* ApcE was assigned. Figure 5C shows the mass spectrum of a trypsin PMF of *G. violaceus* ApcE protein. The coverage of the spectrum was 24% with 24 peptides matched for ApcE. Fortunately, the trypsin peptide at the C-terminus was present in the matched peptides (Panel D), indicating the presence of the complete amino acid sequence of ApcE predicted in the genome and thus justifying the L_{CM}^{129} nomenclature. Finding this peptide in the mass spectrum eliminates the possibility of any post-transcriptional modifications in the C-terminal region of the molecule.

Characterization of the multidomain linker proteins

The high molecular weight linkers L_{RC}^{91} and L_{R}^{81} are multidomain proteins that show characteristic sequences of REP domains. Both linkers have three REP domains and these are shown in Fig. 6A, clustered together in four groups aligned by their similarity to the complete sequence of the CpcC1 (the phycocyanin associated linker) of *G. violaceus*. The sequence alignment clearly shows the strong similarity of these REP domains to CpcC1 and among themselves. They exhibit a pattern composed of three tandem repeats having a coding capacity of 190 amino acid residues in CpcC1. The amino acid sequence of L_{RC}^{91} is composed of an N-terminal segment of 87 amino acid residues followed by rod linker-like REP domains. The REP1 domain stretches from amino acids 88 to 286, REP2 from amino acids 353 to 535 and REP3 from amino acids 623 to 823. The spacing sequences between REP domains are known as Arms. The spacing Arms are 67 and 87 amino acid residues long. The amino acid sequence of L_{R}^{81} is composed of an N-terminal segment of 71 residues, followed by three REP domains and two Arms. The REP domain 1 extends from amino acids 72 to 264, REP domain 2 from amino acids 296 to 493 and REP domain 3 from amino acids 525 to 724. The spacing Arms are very short, both are 32 amino acid residues long (lines 4 and 5 of the first alignment cluster). Note that the Arms of the top cluster from

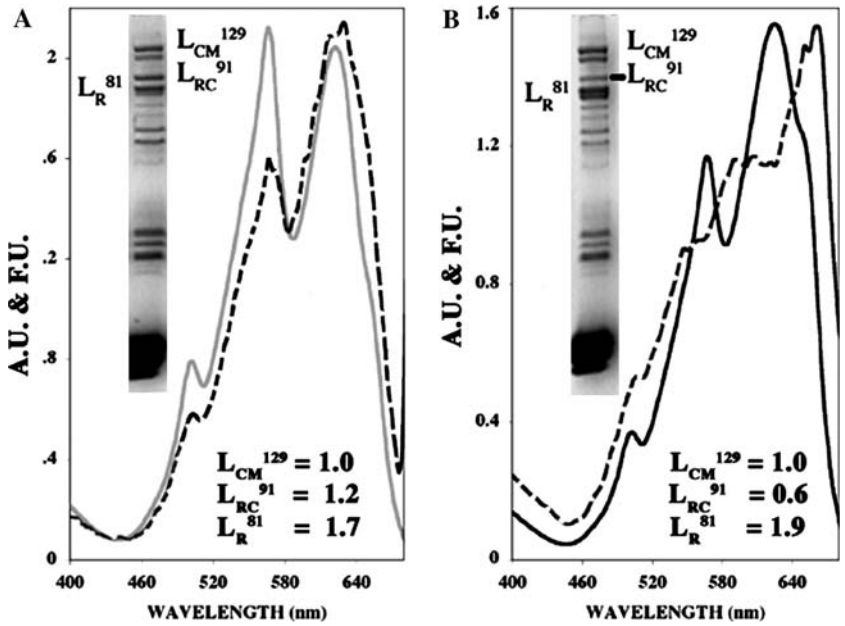

Fig. 4 PBS and PBS dissociation products separated by sucrose gradient centrifugation from Triton X-100 solubilized membranes from *G. violaceus.* Panel **A**: Absorption (solid line) and fluorescence excitation (dashed line) spectra of the PBS. Panel **B**: Absorption (solid line) and fluorescence excitation (dashed line) spectra of the blue band equivalent to band II of Fig. 2D. The relative fluorescence units (F.U.) were normalized with the highest value of absorbance. Inserted in panels **A** and **B** are the electrophoretograms (10% acrylamide) of the corresponding fractions. Underneath the spectra are the identified high molecular weight linkers. Superscripts indicate the molecular masses predicted by amino acid sequences and confirmed by comparison with the mass of the L_{CM}^{99} of *Synechococcus* sp. PCC 7002 (see Fig. 5). The numbers after the equal symbols (=) are the relative stoichiometries of the multidomain linkers obtained from the densitometric analyses of the gels

both linkers differ in length and lack similarity to one another. Figure 6 shows the sequences of $L_{RC}^{91}$and L_R^{81} compared to the rod linker, CpcC1. The amino acid sequences of the first two multi-domain products of 823 and 759 amino acid residues predict polypeptides of masses of 91,780 and 81,350 Da. The molecular mass of ApcE obtained by SDS-PAGE analysis (Fig. 5A, B) was approximately 120 kDa and after trypsin PMF was, more precisely, 129 kDa (Fig. 5C, D). The amino acid sequence of L_{CM}^{129} is composed of 1,155 amino acids and the predicted molecular mass is 129,630 Da. Figure 5D shows the AP subunit domain of L_{CM}^{129}, interrupted by a loop. This interrupted phycobiliprotein is the final energy acceptor of the PBS. It is attached to the multidomain linker (black on white). This multidomain region has four REP domains each approximately 125 amino acids long, homologous to the C-terminal part of CpcC1. The Arms between these domains are longer (87–125 amino acids long) than those seen in the other multidomain linkers (Fig. 6). Based on the presence of these four REP domains in L_{CM}^{129}, we can suggest that *G. violaceus* L_{CM}^{129} might bind four additional trimers in the core of the PBS. Therefore, the PBS of this cyanobacterium might have a pentacylindrical core.

Discussion

G. violaceus is a cyanobacterium that produces PE containing urobilin (Bryant et al. 1981). PE-producing cyanobacteria were classified into three groups based on their ability to adapt physiologically to the color of available light (Bryant 1981; Tandeau de Marsac 1977). In group I, the PBS is independent of the color of light supplied for growth as is the case of *G. violaceus* (Rippka et al. 1974). In group II, PE production is greater in cells grown in green light than in cells that are grown in red light but the level of PC is not responsive to the color of light, as is the case of *Synechococcus marinus* WH8102 (Palenik et al. 2003; Six et al. 2005). In group III, PE production increases in green light, and the level of PC decreases, and in red light only PC is present as in *F. diplosiphon* (Bennett and Bogorad 1973; Tandeau de Marsac et al. 1988). Recently a group IV of chromatic adaptation has been described in some marine *Synechococci* (Everroud et al. 2006). In group IV the PE proteins are the same in different light conditions; the chromophorylation is affected, the phycourobilin and phycoerythrobilin ratio is modified. It has been stated that in *F. diplosiphon*, the CpeR is an activator required for the expression of

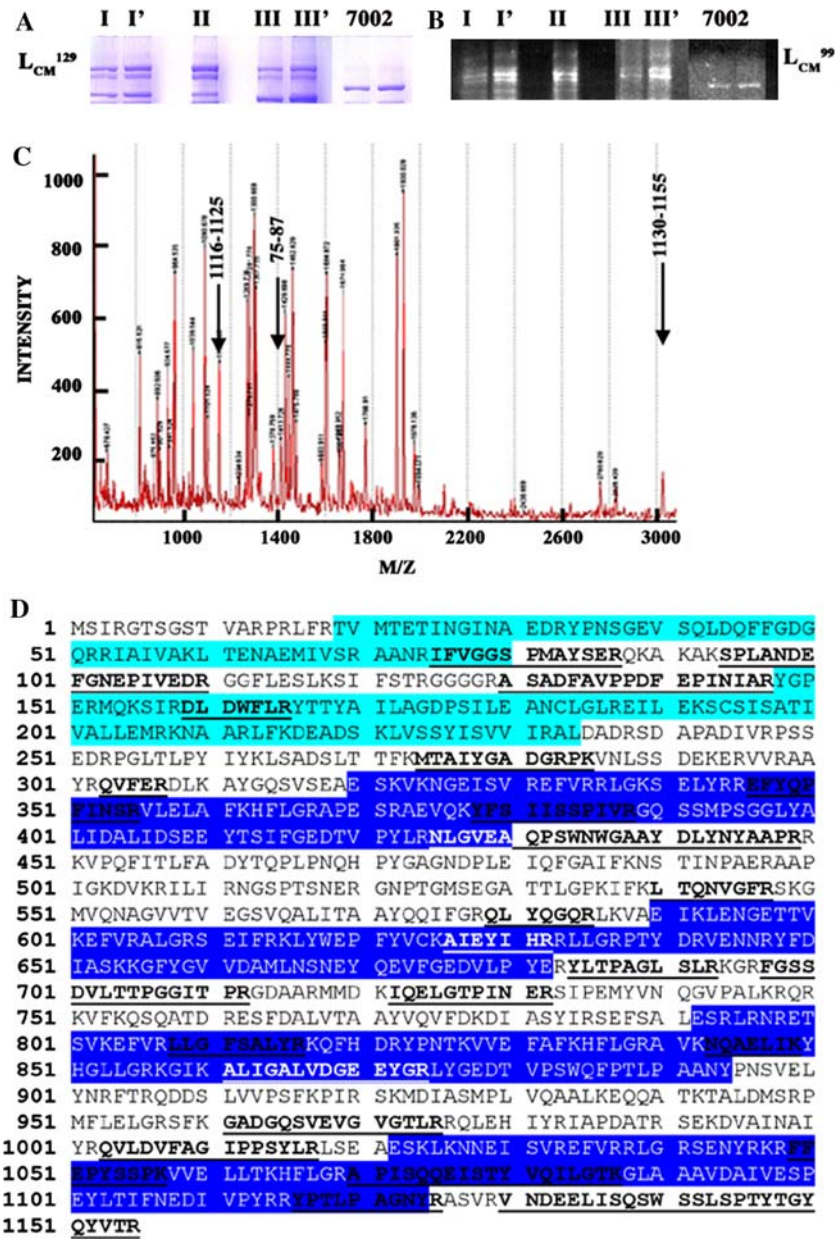

Fig. 5 Panel **A**: Coomassie staining of the LiDS-PAGE (10% acrylamide) of the colored bands from the gradient analyzed in Fig. 4 and compared to the PBS of *Synechococcus* sp. PCC 7002. Panel **B**: The same gel under UV light after spraying with zinc acetate. Lanes I and I' have 75 and 125 μg, respectively of the band I (violet); lane II, 100 μg of the band II (blue) and lanes III and III' 75 and 125 μg of protein, respectively of the band III (fuchsia). The lanes marked 7002 are PBS of *Synechococcus* sp. PCC 7002. The identity of ApcE was determined by PMF. Panel **C**: Peptide mass spectrum of ApcE from *G. violaceus.* After the protein was excised from the gel and digested with trypsin, it was analyzed by MALDI-ToF MS. The number of mass values searched is 36; and the number of mass values matched is 24. The sequence coverage is 24%. Panel **D** shows the amino acid sequence of ApcE (L_{CM}^{129}). The N-terminal segments came from the alignment with the sequence of the α^{AP} subunit of *G. violaceus.* Residues in the regions homologous to *G. violaceus* α^{AP} are shown on light blue. This α^{AP}-like domain is interrupted by a loop. L_{CM}^{129} shows four REP domains in tandem defined after comparison with CpcC1 of *G. violaceus*. Residues of REP domains 1–4, are shown on dark blue. Between the REP domains there are Arms connecting them. Residues of the loop and the Arms, 1–5 top to bottom, are shown as black on white. The 24 amino acid sequences of peptides matched by Mascot program are underlined

the *cpeBA* operon (Cobley et al. 2002). Both the nature and order of genes in *G. violaceus* are different from the two cyanobacterial species belonging to groups II and III in which the genes are clustered in the order *cpeESTR* and are encoded on the same strand. Figure 1B shows that the *G. violaceus cpeR* gene

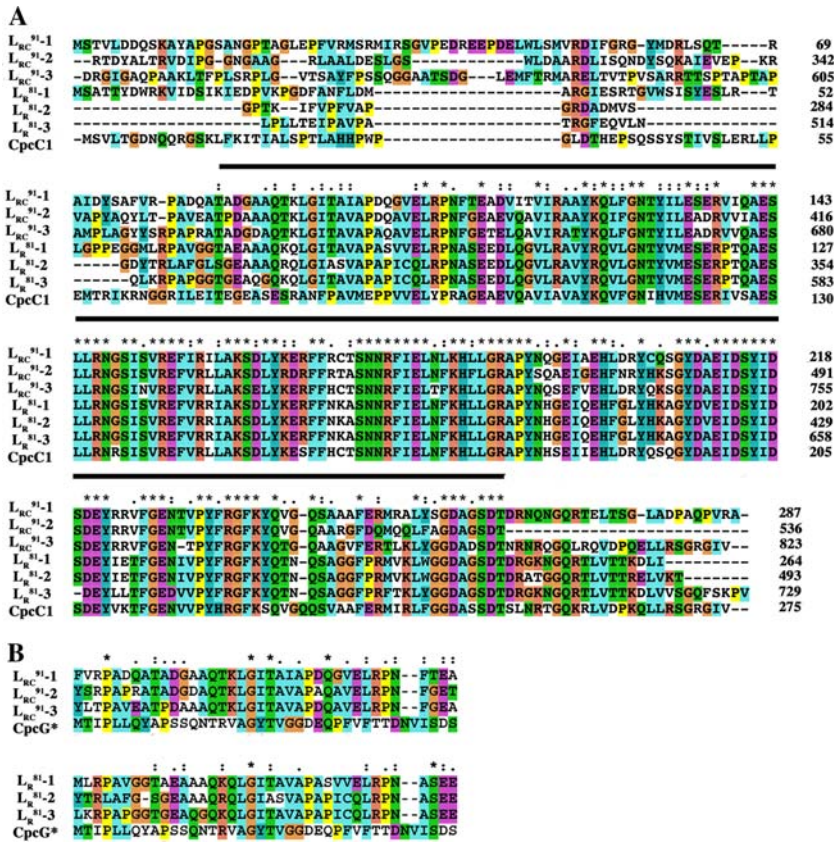

Fig. 6 Amino acid sequence alignment of the REP domains of the multidomain proteins L_{RC}^{91} (Glr1262) and L_R^{81} (Glr2806) and the CpcC1 (Glr0980) linker protein of *G. violaceus* in Panel **A**, the comparison is between the complete sequence of CpcC1, (seventh line at the bottom of each cluster), with the three REP domains of L_{RC}^{91}, (first three lines in each cluster) and with the three REP domains of L_R^{81}, (fourth to sixth lines) in each cluster. Numbers 1–3 indicate the REP domain number in each linker. The top cluster and the first 15 amino acid residues of the second cluster show the Arms sequences of the three REP domains of each multidomain protein and the CpcD-like domain of the CpcC1 protein (amino acids 1–70 of CpcC1). The black bar above the clustered sequences shows the common domain of the alignments. The sequences shown are derived from the complete genome of *G. violaceus* (Nakamura et al. 2003a) and the N-terminal extension of CpcC1 from (Gutiérrez-Cirlos et al. 2006) and are aligned using ClustalX with manual adjustments. Identical residues are marked with an asterisk. The period and colon (./:) indicate sequence similarity. Panel **B**: Sequence alignment of the 40 N-terminal amino acids of the three repeated elements (REP domains 1–3) of the multidomain protein L_{RC}^{91} (Glr1262) with CpcG linker protein of *Synechococcus* sp. PCC 7002 and the REP domains 1–3 of the multidomain linker L_R^{81} (Glr2806) with the same CpcG linker protein of *Synechococcus* sp. PCC 7002

follows the operon *cpcB1A1* and that the phycoerythrin related genes (*cpeBA*, *cpeY* and *cpcZ*) separate it from genes *cpeS* and *cpeT* in the complementary strand. Another peculiarity in the arrangement of the genes that should be noted is the localization of the PC-associated linkers. The *cpcC1–2* genes are separated and the *cpcD1–2* genes are not in the same operon of *cpcBA* as in other cyanobacteria (de Lorimier et al. 1990; Kaneko et al. 1996, 2001; Tandeau de Marsac et al. 1988). Instead *cpcD1–2* seems to be part of the same operon where the PE-associated linkers are encoded. In this work we present evidence that both the two *cpcC*s and the two *cpcD*s of *G. violaceus* are expressed and their protein products are present in the PBS. In addition, two genes with unknown functions were found in the *G. violaceus* genome, *glr1262* and *glr2806* (Nakamura et al. 2003a). The gene *glr1262* is situated in the cluster of genes related to PE and to PC (*glr1260* to *glr1269*): *pebA-pebB-glr1262-cpeC-cpeD-cpeE-cpcD1-cpcD2-cpcE-cpcF* (Fig. 1C). The other novel gene *glr2806* is isolated in the genome as indicated in Fig. 1D. In this work we identified the protein products of these genes as multidomain linkers and assigned them a role in the assembly of *G. violaceus* PBS.

PBSs of the cyanobacterium *G. violaceus* differ from hemidiscoidal PBSs, in three aspects. First, the PBS appearance under the electron microscope shows the rods of the PBS are grouped as a bundle of 6 rods (Guglielmi et al. 1981) giving the PBS a different shape than the typical fan-like rods in the hemidiscoidal PBS (Bryant et al. 1979; Glazer 1989; Grossman et al. 2001; Adir 2005). Second, is the absence of the regular rod-core linkers encoded by the *cpc*G genes (Nakamura et al. 2003a, b). Third, is the presence of two multi-domain linkers found in silico by Nakamura et al. (2003a, b) unique to this cyanobacterium, and experimentally detected as expressed proteins in this article, which are probably responsible for the unique shape of the PBS and the lack of orthodox linkers.

We have constructed a model that reflects both the pigment and linker protein composition and, which includes the characterization of the three multi-domain linkers, which seem to determine the shape of the PBS of *G. violaceus* (Fig. 7). To construct the model we have taken into account the overall morphology of the PBS (Guglielmi et al. 1981) and we have assumed that all the rods are homogeneous in PBPs and in linker composition.

The core

The core membrane linker phycobiliprotein of *G. violaceus* has a molecular mass of 129 kDa (L_{CM}^{129}) as shown in Fig. 5. The polypeptide has four REP domains (Fig. 5) as in *Anabaena* sp. PCC7120 (Ducret et al. 1998), *Anabaena variabilis* (Isono and Katoh 1987) and *Mastigocladus laminosus* (Glauser et al. 1992 and Sidler 1994) that are all members of the so-called pentacylindrical subfamily. The structural implication for the presence of four REP domains in L_{CM}^{129} is a core with 16 AP trimers since there are two L_{CM}^{129} in the core and each REP domain interacts specifically with two AP trimers. The two basal cylinders are organized by: the PBP domain of L_{CM}^{129} (contained in trimer 3 of AP Fig. 7D), the Arm1, REP domain 1 and REP domain 2 (Capuano et al. 1991). The REP domain 3 of each L_{CM} is responsible for binding each hemi-cylinder of the upper complete cylinder. It has been argued that in order to assemble the upper cylinder it is necessary to have antiparallel cylinders in the base (Lundell and Glazer 1983). By the same token it has been asserted that the two hemi-cylinders organized by the REP domain 4 remain as hemi-cylinders and each one must be located on both sides of the upper cylinder (Ducret et al. 1998). However, it has been stated that the PBS of cyanobacteria with an L_{CM} of four REP domains could exhibit tetracylindrical core morphology if their L_{CM}s were oriented in parallel (Anderson and Eiserling 1986). The sequence similarity among L_{CM}s of the different cyanobacteria that have a pentacylindrical core is high, in both their domains, and their arms.

The phycobiliproteins

Since we found evidence both in silico and in experiments (Fig. 5) that L_{CM}^{129}, which constitutes the organizer of the core contains four REP domains, we propose (Fig. 7 Panels C, D and E) a core for the PBS of *G. violaceus* of 16 allophycocyanin trimers. The core structure proposed for this number of allophycocyanin trimers is a pentacylindrical model composed of 3 complete cylinders (4 trimers of AP each) and two hemi-cylinders (H) (two AP trimers) that flank the upper cylinder (Capuano et al. 1991; Sidler 1994). The reconstitution of the core of *M. laminosus* and *Anabaena* sp. PCC 7120 PBS supports this core model (Ducret et al. 1998). The three pigments PE, PC and AP are constitutively expressed in *G. violaceus*. Since there is the equivalent of 8 hexamers of AP in the core, the total number of PCs will be 24 if the PC to AP ratio is 3 and since the PC to PE is 1.33 and the PE to AP ratio is 2.25, there are 18 PEs in six rods. The proportion of the pigments vary between 1 AP per 2.2 to 3.8 PC and 1.7 to 3 per PE, with mean values of 1 to 3 to 2.35, which is close to that predicted by the model. The only reported molar ratio of isolated *G. violaceus* phycobilisomes is 1 to 3.9 to 2.9 for AP to PC and PE, respectively (Guglielmi et al. 1981) for which the model will be 8 hexamers of AP in the core and the rod modified to 5 hexamers of PC and 4 hexamers of PE per rod. The absorbance ratio at 566 nm : 620 nm of the whole cell extract was found to be 2, which is in the limits of the model.

The rods

We have shown that 11 protein linkers are expressed in the PBS of *G. violaceus* (Table 1): two for the core, the L_{CM}^{129} and $L_C^{7.7}$; two novel multi-domain linkers of high molecular weights—91 and 81 kDa; two small linkers ($L_R^{8.1}$ and $L_R^{7.7}$) to cap the rods (this publication) and finally five medium size linkers (Gutiérrez-Cirlos et al. 2006), two PC associated linkers ($L_R^{30.8}$ and L_R^{31}) and three PE associated linkers ($L_R^{31.8}$, $L_R^{28.4}$ and $L_R^{28.3}$). Suggestive evidence concerning the location of the various linkers within the phycobilisome is obtained from the examination of the data in Tables 2, 3 and 4. Three PE disks per rod are needed to fit the expressed three PE-linkers; this proportion is

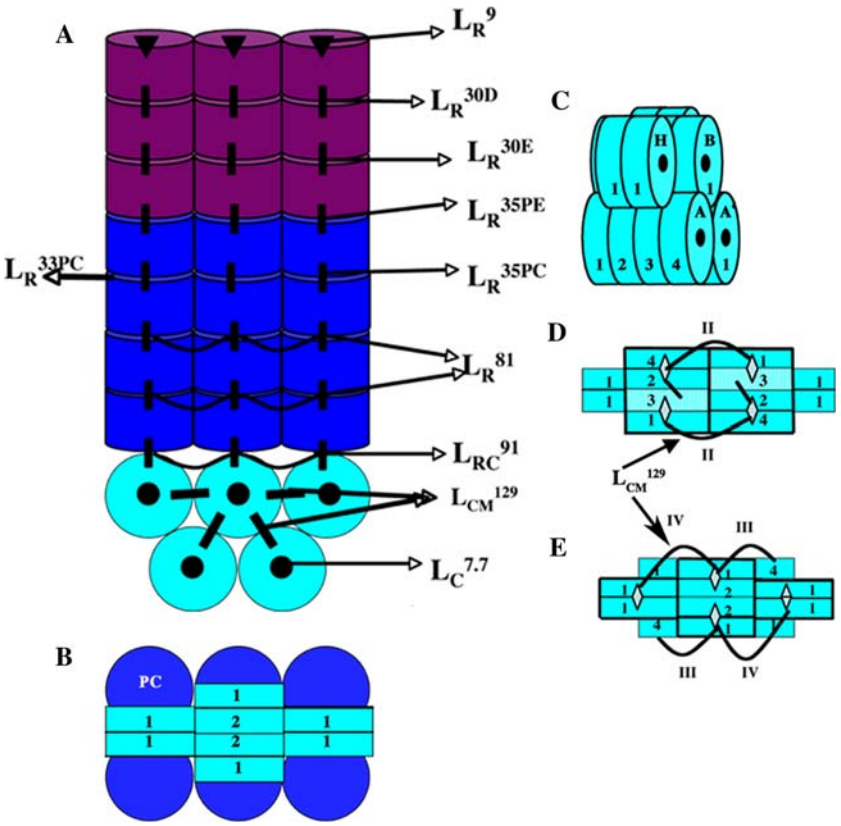

Fig. 7 Schematic representation of the proposed model for localization of the multidomain linkers of the phycobilisome of *G. violaceus*. Light blue, dark blue and purple discs represent trimers of allophycocyanin, hexamers of phycocyanin and hexamers of phycoerythrin, respectively. The single disc in the rods represents hexamers of phycobiliproteins ($\alpha_6\beta_6$). Image **A**. Six peripheral rods are bound as a bundle to the allophycocyanin core. Three long (seven discs) vertical cylinders in the front and three in the back and beneath it are the ends of five horizontal cylinders (three have four trimers of AP, the basal cylinders and the center in the upper level and two, label H, with only two trimers of AP) With respect to the interactions of the 91 kDa domains, we propose that each polypeptide will link three rods with the core taking the place of L_{RC} that is missing in the *G. violaceus* genome. We propose that the 81 kDa polypeptide links six discs of hexameric phycocyanin. It is clear that the presence of these multidomain linkers with short Arms will limit the separation of rods and can explain the parallel rods of the bundle. In image **B**, the PBS shows one cylinder and two half-cylinders (centered) of the second level of the core, which is light blue in color and in the next plane, the six rods are seen in dark blue color. Image **C** shows the arrangement of the AP horizontal cylinders, two basal cylinders and one complete cylinder form a pyramid centered under the vertical rods and two additional half-cylinders (H) in the second level, 16 AP trimers conform with the two L_{CM}^{129}, which maintain their four REP domains and each REP ($\diamond$) is composed of two homologous halves each of which interacts with adjacent trimers (Capuano et al. 1991). Image **D** is the bottom view of the pentacylindrical core. The Roman number II indicates the Arm II, which connects REP domain 1 with REP domain 2. Image **E** is the top view of the pentacylindrical core. The Roman numbers indicate Arms III and IV that connect REP domain 2 with REP domain 3 and REP domain 3 with REP domain 4, respectively. Both half-cylinders (H) are centered to give surface to all of the rods. In the original model, those half-cylinders are at the edges of the complete cylinder (Ducret et al. 1998). The core trimer composition is as follows: 1. $-(\alpha\beta)_3^{AP}-L_C$ 2. $-(\alpha\beta)_3^{AP}$ 3. $-(\alpha_2\beta_3)^{AP}L_{CM}$ 4. $-(\alpha_2\beta_3)^{AP}\alpha^{APB}-L_C$

compatible with our estimates of the relative content of the PE and PE associated linkers in Table 2 ($L_R^{31.8}$, $L_R^{28.4}$ and $L_R^{28.3}$). Interestingly, sequence similarity was found between rod linker polypeptides themselves and between rod linkers and CpcD1 (rod terminator), in the 70 N-terminal amino acids of CpcCs and in the 70 C-terminal amino acids of CpeC. Also in the short alignment (30 amino acids), similarity was found between CpcD2 and CpeD (Table 3). We have hypothesized that the rod linker that attaches the distal PC has a preserved CpcD-like domain and this is a signal or structure that recognizes a CpcD polypeptide or a CpcD like domain. These cap linkers have affinity with a distal PC in the rod, which is attached by the medium size linker that has a CpcD-like domain. The argument is based on an experimental observation using the enzyme ferredoxin-$NADP^+$ oxidoreductase of three domains (FNR-3D). The FNR-3D has a CpcD-like domain that attaches this enzyme to the distal PC in the phycobilisomes and this position is

Table 3 Amino acid sequence identity and similarity values for related linkers

	CpcC1 Glr0950	CpcC2 Gll3219	CpcD1 Gsr1266	CpcD2 Gsr1267	CpeC Glr1263	CpeD Glr1264	CpeE Glr1265
CpcC1 Glr0950		276 aa	n.s.a.	70 aa	178 aa 70 aa	159 aa	149 aa
CpcC2 Gll3219	67/78		n.s.a.	68 aa	159 aa 69 aa	149 aa	159 aa
CpcD1 Gsr1266	n.s.a.	n.s.a.		41 aa	n.s.a.	34 aa	n.s.a.
CpcD2 Gsr1267	52/66	45/61	41/55		70 aa	n.s.a.	n.s.a.
CpeC Glr1263	49/64 52/66	55/66 50/70	n.s.a.	52/69		288 aa	264 aa
CpeD Glr1264	59/73	57/69	41/58	n.s.a.	44/56		242 aa
CpeE Glr1265	57/68	59/73	n.s.a.	n.s.a.	39/56	55/71	

The first horizontal line at the top and the first vertical column at the left hand margin of the matrix identify the seven linker proteins to be compared. The numbers followed by aa are the numbers of amino acids aligned in the comparison. The numbers shown as paired in a ratio are the % identity above and % similarity below. The n.s.a. means no significant alignment. All sequences are from the complete genome of *G. violaceus* (Nakamura et al. 2003a). The CpcC sequences include the N-terminal extensions (Gutiérrez-Cirlos et al. 2006)

preferred over free PC disks or proximal disks (Gómez-Lojero et al. 2003; Schluchter and Bryant 1992). We propose that the position for $L_R^{31.8}$ (CpeC) is in the interface between the distal PC disk and the first PE hexamer. The position of the CpcCs is between the third and the fourth PC. The small length showing similarity between CpcD2 and CpeD suggests that this linker position in the distal PE would attract the CpcD rod terminator at the end of the rod. Finally by exclusion, since the C-terminal of CpeE has no similarity to CpcD, we positioned it to attach the first PE on the second PE disk.

Multidomain linkers: 81 kDa multidomain protein

To understand the composition and localization of the proteins of the *G. violaceus* PBS, we analyzed the colored bands obtained from the Triton X 100 released PBS components and fractionated by sucrose gradient centrifugation (Fig. 2). These bands include the PBS and PBS substructures. Data derived from the heavy blue band (Fig. 2D), with an estimated molecular mass between 1500 and 2000 kDa showed that this band is rich in AP and PC (fraction 5 Fig. 2A, C spectrum 5 and band II Fig. 2D). The presence of core linkers L_{CM}^{129} and the $L_C^{7.9}$ and the presence of an 81 kDa protein in higher concentration (Fig. 3 lane II and Table 2) allows us to conclude that this band contained the core substructure and a rod substructure of 6 PC hexamers organized by the 81 kDa protein linker (L_R^{81}).

The rod core linker (L_{RC}) and the 91 kDa multidomain protein

To get further insight into the localization of the different linkers, we measured the fluorescence excitation spectra of fraction I (PBS) and fraction II (heavy blue band). There was no significant transfer of energy between PC and AP in this fraction (Fig. 4B). In contrast energy transfer does occur in the intact PBS (Fig. 4A). The absence of the 91 kDa proteins from all substructures suggests that the product of *glr1262* gene binds the rods to the core (L_{RC}) and that this binding is released during the dissociation of the PBS. In addition we know from the literature that no sequence similarity was found between rod linkers and the rod core linker polypeptides in the 40 N-terminal amino acids. However, all characterized rod-core linker polypeptides are highly similar in their N-terminal amino acid sequences (Glauser et al. 1992). For example, if the REP domains of 91 kDa protein are aligned to the first 40 amino acids of CpcG of *Synechococcus* sp. PCC 7002, a 45% similarity is found while a similar alignment with the polypeptide of 81 kDa yields only 30% similarity (Fig. 6 Panel B).

The structure as a bundle of rods and the multidomain linkers

The REP domains of these multidomain linkers have two relevant features: an unusual length and the high sequence identity and similarity to REP domains of the

Table 4 Amino acid sequence identity and similarity values of the REP domains of multidomain linkers

	L_{CM}^{129} REP-1	L_{CM}^{129} REP-2	L_{CM}^{129} REP-3	L_{CM}^{129} REP-4	L_{RC}^{91} REP-1	L_{RC}^{91} REP-2	L_{RC}^{91} REP-3	L_{R}^{81} REP-1	L_{R}^{81} REP-2	L_{R}^{81} REP-3
L_{CM}^{129} REP-1	100									
L_{CM}^{129} REP-2	**38/59**	100								
L_{CM}^{129} REP-3	**46/67**	**35/58**	100							
L_{CM}^{129} REP-4	**47/64**	**40/57**	**42/57**	100						
L_{RC}^{91} REP-1	39/56	33/60	34/54	37/58	100					
L_{RC}^{91} REP-2	46/50	37/59	42/58	40/55	**69/81**	100				
L_{RC}^{91} REP-3	38/52	36/58	38/57	37/55	**60/68**	**78/84**	100			
L_{R}^{81} REP-1	42/58	31/58	35/51	38/58	61/71	62/73	64/74	100		
L_{R}^{81} REP-2	42/58	31/58	35/51	39/59	65/75	65/77	58/73	**89/92**	100	
L_{R}^{81} REP-3	43/58	27/48	37/52	40/60	68/75	65/77	60/73	**89/92**	**83/87**	100

Comparison of the amino acid sequences of the linker REP domains in *G. violaceus* phycobilisomes. The first horizontal line at the top and the first vertical column at the left hand margin of the matrix identify the 10 REP domains to be compared. The numbers shown as paired in a ratio are the % identity above and % similarity below. All sequences are from the complete genome of *G. violaceus*. After the alignment of the L_{CM}^{129}, L_{RC}^{91} and L_{R}^{81} with CpcC1 of *G. violaceus*, the REP domains were defined. There are four REP domains for L_{CM}^{129}, three for L_{RC}^{91} and for L_{R}^{81}. Each one was used in a Blastp comparison to the other sequences to obtain the values of identity and similarity presented in the Table

same polypeptide. The sequence length of the REP domains (around 190 amino acids, Fig. 6) is enough to allow them to join hexamers rather than trimers. In contrast, the length of the core membrane linker (L_{CM}) REP domains is around 125 amino acids (Fig. 5D) and it binds trimers (Capuano et al. 1991). The shortness of these Arms will determine the space between different domains and will tie together the rods in the bundle structure of the *G. violaceus* PBS. In addition, the high identity and similarity between REP domains of the same polypeptide (see Table 4, bold numbers) indicate there is interaction between REP domains and so the PBPs must be identical and this also can be expected from the modification and recognition of the PBPs in parallel homogenous disks. L_{R}^{81} has the most homogeneous REP domains, followed in homogeneity by the REP domains of L_{RC}^{91}. Both are clearly different from L_{CM}^{129}, which has heterogeneous REP domains, as might be expected, since the position of these REP domains in the core is in contact with heterogeneous trimers of AP (Fig. 7). Thus the heterogeneity is derived from the different α subunits or by the linker attached to each trimer.

We have built a comprehensive model for the arrangement of pigments and proteins in the PBS light-harvesting complex of the cyanobacterium *G. violaceus*. Our work shows that the complete genome sequences predict differences in the composition of large macromolecular pigment protein assemblies such as the PBS, and that these changes are amenable to characterization by combined spectroscopic and biochemical analyses. Also we have demonstrated that subtle variations in the repetitive sequences of the linker components may give rise to large overall changes in PBS morphology.

Acknowledgements We are greatly indebted to Diego González-Halphen for his editorial assistance and stimulating discussions. We also thank Mary Bower of Purdue University for her expert work in N-terminal sequencing, Jorge Zarco Mendoza for his technical assistance and Cristina Díaz-Trueba for her secretarial assistance. We also appreciate critical reading of the manuscript by Loretta Krogmann and Michele Krogmann.Grant SEP-CONACYT-México supported this study. Reference Number 46416-Q.

References

Adir N (2005) Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. Phosynth Res 85:15–32

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped Blast and PSI Blast: a new generation of protein database search programs. Nucleic Acid Res 25:3389–3402

Anderson L, Eiserling FA (1986) Asymmetrical core structure in phycobilisomes of the cyanobacterium *Synechocystis* sp. PCC 6701. J Mol Biol 191:441–451

Bennett A, Bogorad L (1973) Complementary chromatic adaptation in a filamentous blue green alga. J Cell Biol 58:419–435

Bryant DA (1981) The photoregulated expression of multiple phycocyanin species. Eur J Biochem 119:425–429

Bryant DA, Guglielmi G, Tandeau de Marsac N, Castets AM, Cohen-Bazire G (1979) The structure of cyanobacterial phycobilisomes: a model. Arch Microbiol 123:113–127

Bryant DA, Cohen-Bazire G, Glazer AN (1981) Characterization of the biliproteins of *Gloeobacter violaceus* chromophore content of a cyanobacterial phycoerythrin carrying phycourobilin chromophore. Arch Microbiol 129:190–198

Bryant DA, Stirewalt VL, Glauser M, Frank G, Sidler W, Zuber H (1991) A small multigene family encodes the rod-core linker polypeptides of *Anabaena* sp. PCC 7120 phycobilisomes. Gene 107:91–99

Capuano V, Braux AS, Tandeau de Marsac N, Houmard J (1991) The anchor polypeptide of cyanobacterial phycobilisomes. J Biol Chem 266:7239–7247

Cobley JG, Clark AC, Weerasurya S, Queseda FA, Xiao JY, Bandrapali N, D′Silva I, Thounaojam M, Oda JF, Sumiyoshi T, Chu MH (2002) CpeR is an activator required for expression of the phycoerythrin operon (cpeBA) in the cyanobacterium *Fremyella diplosiphon* and is encoded in the phycoerythrin linker-polypeptide operon (cpeCDESTR). Mol Microbiol 44:1517–1531

Delepelaire P, Chua NH (1979) Lithium dodecyl sulfate, polyacrilamide gel electrophoresis of thylakoid membranes at 4°C; characterizations of two additional chlorophyll a protein complexes. Proc Natl Acad Sci USA 76:111–115

de Lorimier R, Bryant DA, Stevens SE (1990) Genetic analysis of a 9 kDa phycocyanin-associated linker polypeptide. Biochim Biophys Acta 1019:29–41

Ducret A, Müller SA, Goldie KN, Hefti A, Sidler WA, Zuber H, Engel A (1998) Reconstitution, characterization and mass analysis of the pentacylindrical allophycocyanin core complex from the cyanobacterium *Anabaena* sp. PCC 7120. J Mol Biol 278:369–388

Everroad C, Six C, Partensky F, Tomas JC, Holtzendorff J, Wood AM (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. J Bacteriol 188:3345–3356

Füglistaller P, Mimuro M, Suter F, Zuber H (1987) Allophycocyanin complexes of the phycobilisome from *Mastigocladus laminosus.* Influence of the linker polypeptide $L_c^{8.9}$ on the spectral properties of the phycobiliprotein subunits. Biol Chem Hoppe Seyler 368:353–367

Glauser M, Bryant DA, Frank G, Wehrli E, Rusconi SS, Sidler W, Zuber H (1992) Phycobilisome structure in the cyanobacteria *Mastigocladus laminosus* and *Anabaena* sp. PCC 7120. Eur J Biochem 205:907–915

Glazer AN (1989) Light Guides. J Biol Chem 264:1–4

Gómez-Lojero C, Pérez-Gómez B, Prado-Flores G, Krogmann DW, Cárabez-Trejo A, Peña-Díaz A (1997) The phycobilisomes of the cyanobacterium *Arthrospira* (*Spirulina*) *maxima*. Int J Biochem Cell Biol 2910:1191–1205

Gómez-Lojero C, Pérez-Gómez B, Shen G, Schluchter WM, Bryant DA (2003) Interaction of ferredoxin: $NADP^+$ oxidoreductase with phycobilisomes and phycobilisome substructures of the cyanobacterium *Synechococcus* sp. strain PCC 7002. Biochemistry 42:13800–13811

Gottschalk L, Fischer R, Lottspeich F, Scheer H (1991) Origin of the red shifted absorption in phycocyanin 632 from *Mastigocladus laminosus*. Photochem Photobiol 54:283–288

Gottschalk L, Lottspeich F, Scheer H (1993) Reconstitution of allophycocyanin from *Mastigocladus laminosus* with isolated linker polypeptide. Photochem Photobiol 58:761–767

Grossman AR (2003) A molecular understanding of complementary chromatic adaptation. Photosynth Res 76:207–215

Grossman AR, Bhaya D, He Q (2001) Tracking the light environment by cyanobacteria and the dynamic nature of light harvesting. J Biol Chem 276:11449–11452

Guglielmi G, Cohen-Bazire G, Bryan DA (1981) The Structure of *Gloeobacter violaceus* and its phycobilisomes. Arch Microbiol 129:181–189

Gutiérrez-Cirlos EB, Pérez-Gómez B, Krogmann DW, Gómez-Lojero C (2006) The phycocyanin-associated rod linker proteins of the phycobilisome of *Gloeobacter violaceus* PCC 7421 contain unusually located rod-capping domains. Biochim Biophys Acta 1757:130–134

Isono T, Katoh T (1987) Subparticles of *Anabaena* phycobilisomes II. Molecular assembly of allophycocyanin cores in reference to ''anchor'' protein. Arch Biochem Biophys 256:317–324

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3:109–136

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M, Tabata S (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 8:205–213

Liu LN, Chen XL, Zhang YZ, Zhou BC (2005) Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. Biochim Biophys Acta 1708:133–142

Lundell DJ, Glazer AN (1983) Molecular architecture of a light-harvesting antenna. Core substructure in *Synechococcus* 6301 phycobilisomes two new allophycocyanin and allophycocyanin B complexes. J Biol Chem 258:902–908

MacColl R (1998) Cyanobacterial phycobilisomes. J Struct Biol 124:311–334

Martin RG, Ames BN (1961) A method for determining the sedimentation behavior of enzymes: application to protein mixtures. J Biol Chem 236:1372–1378

Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa Ch, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi Ch, Yamada M, Tabata S (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res 9:123–130

Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa Ch, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi Ch, Yamada M, Tabata S

(2003a) Complete genome structure of *Gloeobacter violaceus* PCC7421, a cyanobacterium that lacks thylakoids. DNA Res 10:137–145

Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa Ch, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi Ch, Yamada M, Tabata S (2003b) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylacoids (Supplement). DNA Res 10:181–201

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J (2003) The genome of a motile marine *Synechococcus*. Nature 424:1037–1042

Raps S (1990) Differentiation between phycobiliprotein and colorless linker polypeptides by fluorescence in the presence of $ZnSO_4$. Plant Physiol 92:358–362

Reuter W, Wiegand G, Huber R, Than ME (1999) Structural analysis at 2.2 Å of orthorhombic crystals presents the asymmetry of the allophycocyanin-linker complex, $AP.L_c^{7.8}$, from phycobilisomes of *Mastigocladus laminosus*. Proc Natl Acad Sci USA 96:1363–1368

Rippka R, Waterbury J, Cohen-Bazire G (1974) A cyanobacterium which lacks thylakoids. Arch Microbiol 100:419–436

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111:1–61

Schägger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–379

Schluchter WM, Bryant DA (1992) Molecular characterization of ferredoxin-$NADP^+$ oxidoreductase in cyanobacteria: cloning and sequence of the *pet*H gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. Biochemistry 31:3092–3102

Sidler WA (1994) Phycobilisome and phycobiliprotein structures. In: Bryant DA (ed) The molecular biology of cyanobacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 139–216

Six C, Thomas JC, Thion L, Lemoine Y, Zal F, Partensky F (2005) Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. J Bacteriol 187:1685–1694

Tandeau de Marsac N (1977) Ocurrence and nature of chromatic adaptation in cyanobacteria. J Bacteriol 130:82–91

Tandeau de Marsac N (2003) Phycobiliproteins and phycobilisomes: the early observations. Photosynth Res 76:197–205

Tandeau de Marsac N, Mazel D, Damerval T, Guglielmi G, Capuano V, Houmard J (1988) Photoregulation of gene expression in the filamentous cyanobacterium *Calothrix* sp. PCC 7601 light-harvesting complexes and cell differentiation. Photosynth Res 18:99–132

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

van Thor JJ, Gruters OWM, Matthijs HCP, Hellingwerf KJ (1999) Localization and function of ferredoxin: $NADP^+$ reductase bound to the phycobilisomes of *Synechocystis*. EMBO J 18:4128–4136