

Model for the phycobilisome rod with interlocking disks based on domain-weighted linker-polypeptide sequence homologies of *Mastigocladus laminosus*

Axel Parbel and Hugo Scheer

Botanisches Institut, Ludwig-Maximilians Universität, Menzingerstr. 67, D-80638 München, Germany

ABSTRACT. The 8 linker-polypeptides from the cyanobacterium *Mastigocladus laminosus* were examined by comparative amino-acid sequence analysis for the predicted domain structure reported in the literature (Glauser 1991, Esteban 1993) in detail using split sequences for the rod, rod-core and for the repeats from the core-membrane linker-polypeptide ($L_{cm}^{127.6}$). This analysis gives two distinct domains, where domain 1 (~ 22 kDa, identity between 31 and 70%) is present in the N-terminal two thirds of the class II linkers (~ 30 kDa) and in the repeats of the $L_{cm}^{127.6}$, and domain 2 (~ 10 kDa, identity between 28 and 41%) in the C-terminal part of the class II rod linkers ($L_r^{31.5}$ and $L_r^{32.5}$) and in the two capping linkers ($L_c^{7.7}$ and $L_r^{8.2}$). Based on these data, X-ray structure analysis from phycobiliproteins and proteolysis experiments, an interlocking model for the phycobilisome rod organization is proposed, with linkers protruding from one phycobilisome disk into the neighbouring one.

Abbreviations: APC= allophycocyanin (encoded by *apcA* and *apcB*, for the α - and β -subunits, respectively), *apc*= gene locus of allophycocyanin, *cpc*= gene locus of phycocyanin, L= linkerpolypeptide [the superscripts denote the molecular weights, according to protein sequence calculation, the subscripts give the location denoted as r (rod), rc (rod-core), c (core), and cm (core-membrane)], the correspondance of genes and the derived linker polypeptides are as follows: $L_r^{8.2}$ (*cpcD*), $L_r^{32.3}$ (*cpcC*), $L_r^{31.5}$ (*pecC*), $L_{rc}^{31.7}$ (*cpcG1*), $L_{rc}^{28.7}$ (*cpcG2*), $L_{rc}^{29.6}$ (*cpcG3*), $L_c^{7.7}$ (*apcC*) and $L_{cm}^{127.6}$ (*apcE*), *M.*= *Mastigocladus*, PBS= phycobilisome, PC= phycocyanin (encoded by *cpcB* and *cpcA*, for the β - and α -subunits), PE= phycoerythrin, PEC= phycoerythrocyanin (encoded by *pecB* and *pecA*, for the β - and α -subunits), *pec*= gene locus from phycoerythrocyanin.

1. INTRODUCTION

Phycobilisomes (PBS) are the extra-membraneous light-harvesting complexes of cyanobacteria, rhodophytes and cyanelles which funnel the light energy absorbed by their 500 to >1000 chromophores to two specialized ones, and from their to the reaction centers, with close to 100% quantum yield (Porter *et al.* 1978, Manadori and Melis 1985). The PBS of the cyanobacterium, *Mastigocladus* (*M.*) *laminosus* (*Fischerella* PCC 7603), contains at least 16 different polypeptides used each in several (up to 108) copies (Sidler *et al.* 1981, Zuber 1986, Glauser 1991, Esteban 1993), other PBS can be even more complex (Gantt 1986). Biochemically, the two major PBS components are the phycobiliproteins carrying covalently bound linear tetrapyrrole chromophores (MacColl and Guard-Friar 1987), and the linker polypeptides which have primarily a structural role, but are involved as well in the fine-tuning of phycobiliprotein absorption and can also carry chromophores (Glazer 1984). Morphologically, the two major components are

(a) a central core comprised of allophycocyanins (APC) and the core linkers ($L_c^{7.7}$, $L_{cm}^{127.6}$), and (b) several rods which in *M. laminosus* are comprised of phycocyanin (PC), phycoerythrocyanin (PEC), and the rod-linkers ($L_r^{8.2}$, $L_r^{31.5}$, $L_r^{32.3}$, $L_{rc}^{31.8}$, $L_{rc}^{28.7}$, and $L_{rc}^{29.6}$). PEC can be replaced by PC depending on the growth conditions (irradiation and CO₂ content, Esteban 1993).

The genome organisation and DNA or amino-acid sequences of all known components of the PBS from *M. laminosus* have been determined (Zuber 1986, Eberlein and Kufer 1990, Glauser *et al.* 1992, Esteban 1993, Kufer, Gene Bank database but otherwise unpublished results). This includes three rod-core linkers, which have been located on the *cpc*-operon, while a fourth one is still under discussion. For several linker-free phycobiliproteins including PC and PEC of *M. laminosus* (Schirmer *et al.* 1986, Schirmer *et al.* 1987, Düring *et al.* 1990, Ficner *et al.* 1992) high-resolution three dimensional structures are available. The doughnut-shaped tri- ($\alpha_3\beta_3$) or hexamers ($\alpha_6\beta_6$) are the basic phycobilisome building blocs. They appear to be at the same time the largest biliprotein aggregates which can be formed in the absence of linker polypeptides (Lundell *et al.* 1981, Fischer *et al.* 1990, Gottschalk *et al.* 1991, Glauser *et al.* 1992). The latter are then mainly responsible for the assembly of the tri- and hexameric biliproteins to the entire phycobilisome, where they must provide both the interaction energies and specificities (Lundell *et al.* 1981, Glazer 1984, Tandeau de Marsac *et al.* 1988).

According to electron microscopy (Mörschel *et al.* 1977) the linkers are located in the interior holes of the tri- and/or hexameric biliproteins. This view has been strengthened by the X-ray structure of a B-PE-hexamer containing a chromophore-bearing γ -subunit believed to replace the linker (Ficner and Huber 1993). Although no details of this γ -subunit are discernible due to the

threefold disorder induced by the C_3 -symmetry of the biliprotein, it fills snugly the central hole of a single hexamer. The X-ray structure has recently become available (Reuter *et al.* 1999) of a small linkerpolypeptide within an AP trimer ($\alpha_3\beta_3 * L_c^{7.7}$) from the phycobilisome of the cyanobacterium, *M. lamosus*. This structure provides several new aspects:

1. The linker, a three-stranded β -sheet and two α -helices, is in contact with only two of the three β -subunits and interacts directly with the corresponding chromophores of the phycobiliprotein subunits.

2. It influences the structure of the $\alpha\beta$ -subcomplexes, bringing closer together the β -chromophores.

3. In spite of the generally acidic character of the phycobiliproteins and basic character of the linker, there is no evident clustering of charged, polar and hydrophobic residues at the protein-protein interface.

4. The linker polypeptide $L_c^{7.7}$ fills only partly the central cavity of the phycobiliprotein trimer.

While the small $L_c^{7.7}$ polypeptide is completely located within the central cavity, and there is no obvious protrusion of the larger γ -subunit from the PE-trimer in B-PE, a different view comes from biochemical evidence for the type 2 (≈ 30 kDa) linkers. Limited (deliberate or nondeliberate) proteolysis experiments of small phycobiliprotein-linker complexes indicate that certain parts of the linkers are readily cleaved, while the remainder is much more resistant. From *M. lamosus*, several such partially digested complexes have been isolated and characterized (Esteban *et al.* 1990, Gottschalk *et al.* 1991). While they do no longer assemble to larger aggregates, they show the characteristic spectral properties of the complexes carrying the respective integral linker polypeptide. Based on this differential digestion, in combination with comparative sequence analysis, Wilbanks and Glazer (1993) have proposed a model for the arrangement of the γ -subunit in a cyanobacterial PE hexamer, in which only the stable N-terminal portion of the linker occupies the central hole, whereas the cleavable C-terminal, amino-acid residues protrude from the central hole.

In the framework of a project aimed at investigating structure-function relationships in linker polypeptides, we have recently resequenced part of the *cpc*-operon of *M. lamosus* and analysed the domain structure of all linkers. Here we wish to extend this idea of part of the linkers protruding from the central hole, such that this leaves part of this hole at the other end of the hexamer empty. This can function as a docking place for the C-terminus of the linker of the neighbouring hexameric disk or, at the end of the rod, with a small capping linker, giving rise to an interlocking model of discs in the rods.

2. MATERIALS AND METHODS

EMBL-Phage bank. *M. lamosus* was grown at 55 °C in Castenholz-medium d (Castenholz 1970). The cells were washed and lysed according to the method of Mazel *et al.* (1996), and the DNA purified by

CsCl/ethidiumbromid gradient-centrifugation as described by Lau *et al.* (1987). Genomic DNA (100 μ g) was partially digested with *Sau*3A, and the fragments fractionated on a 10–40% linear sucrose-gradient in 10 mM Tris-HCl buffer (pH 7.6), containing 1 mM EDTA. Centrifugation was carried out in a swing-out rotor for 24h at 120 000 *g.

The fragments of ~ 16 kb size were ligated into *Bam*HI-cleaved EMBL3 phage vector (Frischauf *et al.* 1983), and the phages grown in *E. coli* strains LE392 and Y1088 (Ausubel *et al.* 1987). The genomic library was screened with a radiolabelled DNA fragment, encoding part of *cpcAC* (radiolabelling was carried out using a nick translation kit, Boehringer Mannheim) and hybridization under stringent conditions according to Maniatis *et al.* (1982).

Subclones were constructed in plasmid vector pUC19 (Yannish-Perron *et al.* 1985) maintained in *E. coli* strain DH5 α (Ausubel *et al.* 1987). Transformations were carried out according to the standard procedure of Hanahan (1985).

DNA sequencing. Plasmid DNA to be used in sequence analysis reactions was isolated by an alkaline lysis method (Heinrich 1986). The sequencing reactions were primed using the pUC/M13 sequencing primers (Boehringer Mannheim) and labelled with 10 μ Ci [α - 32 P] dATP using a T7 DNA sequencing kit (Pharmacia). Samples were electrophoresed on 6 or 8% denaturing polyacrylamid gels, which were subsequently dried on Whatmann 3 MM paper before exposure to X-ray films (Amersham).

Alignment of sequences. Amino acid sequences were either analysed by pairwise alignments using the FASTA program (Pearson and Lipman 1988), or by multiple alignment using the HUSAR system (Deutsches Krebs Forschungs Zentrum, Heidelberg) and CLUSTAL V (Higgins and Sharp 1988). Calculated homologies count unmatched amino acids as mismatches, unless otherwise stated. For definition of the domain boundaries see results and discussion.

Published sequences used in this analysis were taken from Genbank accession number M75599 (Kufer, privat communication) for the genes *pecC*, *cpcC*, and *cpcD*, from Glauser (1991) for *cpcG1*, *cpcG2*, and *cpcG3* and from Esteban (1993) for *apcC* and *apcE*.

3. RESULTS AND DISCUSSION

The work is based on a 16 kbp EMBL phage fragment containing the *pec*-operon and the 3'-adjacent *cpc*-operon. The sequence of *pecB* and *pecA* have been published by Eberlein and Kufer (1990), those of the entire *pec*-operon and *cpcB* through D were deposited in the Gene Bank database (see Material and Methods) but have not yet been published, and that from the end of *cpcF* through *cpcG3* by Glauser (1991). The sequencing strategy and the nucleotide and derived protein sequences for *cpcF* and *cpcG1* are shown in Figure 1.

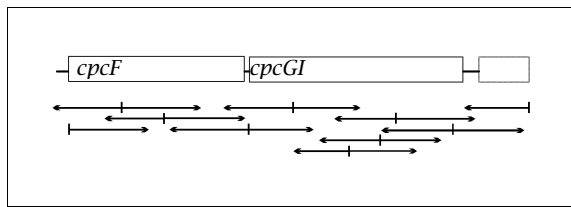


Figure 1. **A**: Sequencing strategy for the sequence shown in **B**. **B** (next page): Nucleotid and derived amino-acid sequence of the sequenced genes *cpcF* and *cpcG1*.

The structure determined for the *cpc*-operon is identical to that published by Glauser *et al.* (1992). *CpcB* and *cpcA* coding for the β - and α -subunits of PC, respectively, are followed by two linker genes (*cpcC* coding for L_r^{32.3} and *cpcD* coding for the rod-terminating L_r^{8.2}, Lundell *et al.* 1981). The next two orfs (*cpcE*, *cpcF*) code for biosynthetic factors involved in the chromophore attachment (Zhao *et al.* 2000). The derived proteins are not yet well known in *M. laminosus*, but their sequences are homologous (44.1% for *pecE*, 49.0% and 28.2% for *cpcF* and *pecF*, data not shown) to those of the two components of the α 84 chromophore ligase from *Synechococcus* sp. PCC 7002 (Fairchild and Glazer 1994).

The last three genes (*cpcG1*, *cpcG2*, *cpcG3*) code for the linkers responsible for attachment of the rods to the core ($L_{rc}^{31.8}$, $L_{rc}^{28.7}$, $L_{rc}^{29.6}$, respectively). The search for a fourth one, which has been suggested by Glauser *et al.* (1992) was negative. Using a homologous N-terminal *cpcG1* probe (nucleotide 758 to 939, Figure 1B) under stringent conditions (using 40% formamid in hybridization buffer at a temperature of 42 °C) an identical hybridization pattern was obtained with four restriction enzymes (*NheI*, *EcoRI*, *PstI*, and *HindIII*), starting from

either the 16 kbp EMBL phage fragment, or from genomic DNA from *M. lamosus*. If there is an additional rod-core linker, it should have a significantly different sequence in the probed region.

The *pec*-operon has a similar ordering (*pecB*, *pecA*, *pecC*, *pecE*, *pecF*), but lacks a D-gene (coding for the small rod linker), and the G-genes coding for the rod-core linkers. PEC is believed to be located at the distal ends (Bryant *et al.* 1979) or at least in the central portion of the rods (Reuter and Nickel-Reuter 1993), the lack of the G-genes is therefore expected. The absence of a small rod linker in the *pec*-operon is surprising if a terminal location is assumed. One might speculate, that in view of the rather similar three-dimensional structures of PC and PEC, $L_r^{8,2}$ can cap either biliproteins at the ends of the rods. Reuter and Nickel-Reuter (1993) have questioned this view and used stoichiometric arguments for positioning (at least some of the) PEC in the center of the rods.

A comparison of the derived amino-acid sequences of the *M. laminosus* linkers (Figure 2) indicates that all of them are composed of a set of domains, which are combined differently for the different linkers (identity values are reported in Table 1). The domains were found by comparing first the class II rod (L_R) to the small class III linkers (L_R and L_C). (The terminology of class I, II, and III is based on their size in the 100, 30, and 9 kDa range, respectively Tandeau de Marsac *et al.* 1988). High homologies with the latter were found in the C-terminal parts of the former, which defines what we would like to call the domain 2. The remaining N-terminal parts of the class II rod linkers were then compared with each other and the L_{RC} and the L_{CM} . This analysis corroborates the results of Glauser (1991) and Esteban (1993) with two alterations:

Table 1. Homologies among the phycobilisome linkers or linker domains. The regions used for comparison are shown in Figure 2, with “N” and “C” denoting the N- and C-terminal parts of the class II linkers (L_r^{31.5}, L_r^{32.5}, L_{rc}^{31.7}, L_{rc}^{28.7}, and L_{rc}^{29.6}), and RA through RD denoting repeats A through D (250 to 410, 550 to 690, 700 to 870, and 950 to 1110) of L_{cm}^{127.6}. The values given in the table are identities of amino acids calculated from the FASTA program, “-” shows that no homologies were indicated by the used programs.

		$L_r^{31.5}$ (<i>pecC</i>)		$L_r^{32.3}$ (<i>cpcC</i>)		$L_{rc}^{31.7}$ (<i>cpcG1</i>)		$L_{rc}^{28.7}$ (<i>cpcG2</i>)		$L_{rc}^{29.6}$ (<i>cpcG3</i>)		$L_r^{8.2}$ (<i>cpcD</i>)	$L_c^{7.7}$ (<i>apcC</i>)	$L_{cm}^{127.6}$ (<i>apcE</i>)			
		N	C	N	C	N	C	N	C	N	C			RA	RB	RC	RD
$L_r^{31.5}$ (<i>pecC</i>)	N	100.0	-	54.3	-	30.9	-	32.9	-	34.1	-	-	-	34.3	40.2	35.9	38.1
	C		100.0	-	39.4	-	-	-	-	-	-	40.3	32.1	-	-	-	-
$L_r^{32.3}$ (<i>cpcC</i>)	N			100.0	-	35.0	-	35.8	-	35.4	-	-	-	33.1	39.4	35.8	33.8
	C				100.0	-	-	-	-	-	-	40.9	28.3	-	-	-	-
$L_{rc}^{31.7}$ (<i>cpcG1</i>)	N					100.0	-	66.0	-	59.1	-	-	-	35.3	38.4	33.8	36.4
	C						100.0	-	-	-	-	-	-	-	-	-	-
$L_{rc}^{28.7}$ (<i>cpcG2</i>)	N							100.0	-	70.2	-	-	-	35.7	33.5	35.2	33.6
	C								100.0	-	-	-	-	-	-	-	-
$L_{rc}^{29.6}$ (<i>cpcG3</i>)	N									100.0	-	-	-	34.3	36.8	30.8	30.3
	C										100.0	-	-	-	-	-	-
$L_r^{8.2}$ (<i>cpcD</i>)												100.0	33.3	-	-	-	-
$L_c^{7.7}$ (<i>apcC</i>)												100.0	-	-	-	-	-
$L_{cm}^{127.6}$ (<i>apcE</i>)	RA													100.0	44.4	43.3	41.5
	RB														100.0	39.2	41.0
	RC															100.0	47.3
	RD																100.0

CTTACCAATTACTGACATATCAACTGCTAACTACTATACTTTTGACCATGACTGCCACTAC	60
M T A T T	
CCAACAATTAATCCGCAGCGTCGCTGAAGCAGACTACCAGCACAATGACAGCAGCTGT	120
Q Q L I R S V A E A D S P A Q M C T A A V	
AAAAAAATTGGCAGCTGCCAAAGATGAAGCAGCAATTCCCACTTTAATTCGCGTATTCGG	180
K K L L A A A K D E A A I P T L I A V F G	
TTATAACAACCCGACAGCAGCAGCAGCAGTGTCTGGGCTCATTTGCATTTAAAGAAGC	240
Y N N P E A A A A A A A K G L I A F K E A	
AGCAGTACCGTATTGTAGTCCCAAATTGACGATTACAACCTACGGTGCAGCAGCTTATTC	300
A V P Y L L A Q I D D Y N Y G A R A Y S	
GATCAGAACTAGCTGCGATTCGCCGATCCCCGCCCTAGAAATTGCTATCATCCGC	360
I R T L A A I A D P R A L E V L L S S A	
AGTAAAGCACTTTGCTCCAGCGTCCGCTGCGCCGCCCAAAGGATTGGGAAACCTGCA	420
V N D F F A G S P N V R C I A A K G L A N L Q	
GTTTTGCAAACCTAGACCCCTCCCCAACGGCAACAGCACAACAAAGCCTTGAAACTTT	480
F C K L D P P Q R Q T A Q T K A L E T L	
GCTGCTTGTCTCCCAAGCCCAATTTGTTAATTCGTTACGCTGCTGTTGTCGGTTTGA	540
L L V S Q A D C P N W L I R Y A A A K G L V G L E	
AGCTCTCGCAAAATACCTCAGCTACAACAACCAATTCAAACAAGATTGCACAAATACT	600
A L A K I P Q L Q Q P I Q T R F A Q I L	
CGCCACTGACACCGAGCAAGCTATTGTCAGCTCGCGTCAACTGGCTCAAAAAGTAGGAA	660
A T D T E Q A A I C S R A T G T S K T R	
TAGAACAACCTACAAACCGAAACTTAAACCAAGCTTTCGTACTCTCTTAATTTTGT	720
R R Q T T N R K L K P A F V L L L I L F	
TGTTCCACTCTCAAGATAAACTAAGTCAAAGTAAAAATGCCAATACCACCTGCTTCCCT	780
V P L S R - M P I P L L S F	
TCCTCTGACCACCCAAACCGAGCTGTCGAAGGCTACGAAGTACCAACGAAGACACACC	840
P L T T Q N Q R V E G Y E V P N E D T P	
AACAATATTGATCGAGATACTTCTTCAGATCAAGAAATCGATGCCATCATTTGGCG	900
T I Y R L T D T S D T E I D A I I W A	
AGCTTATCGACAAATATTGACGGAACCTTAATCTAGAAATACCGCCAACTTACCT	960
A Y R Q I F S E H L I L E T Y R Q P Y L	
AGAATCACAACCTGCAAAACGAGCGATAAACGTCGCGATTTTATTCGCGGATTGGCAA	1020
E S Q L R N R A I N V R D F I R G L G K	
ATCAGAAGTATCCGCAAGAAGTAGCCCAAACAACCTCAACCTACCGTTTGGTTGACAT	1080
S E V Y R Q E V A T S N S N Y R L V D I	
CAGCTTCAAACGCTTTTAGGACGCGCTACCTACGTTAAAGCGAACAATCGCCTGGTC	1140
S F K R F L G R A T Y G K S E Q I A W S	
AATCGTCACCGCAACCCAAAGGATTGACGCGTTTCATCGACGCTCTGGTAGATACGAAGA	1200
I V T A T Q G L H G F I D A L V D S E E	
GTACCGCAAAAACCTTCGGCGAGATATCGTACCTTTCCAACGCCCGCGCTTCAAGGATCG	1260
Y R Q N F G D D I V P F Q R R R F K D R	
TCCTTTAATTAGTCAACCCCGCTACGCCGATTACTGGCGCAATCGCTTCCTAGAAC	1320
P F N L V N P R Y A D Y W R N R L L E Q	
ATTTTATGGCGGCAAGCTTCTTACGAGTCGTCGCGCGGAGAATCAGGAAACGAGG	1380
F L G G Q Y F Y R V R A C C G S G K R G	
AGTTGCTGGCGCAATTCCAGCAGCTTTCTTATCCATGGCCGCAAGCATTGCTCCGACGG	1440
V A G A I P S T F L S M A A A S I A P S G	
AATCAGCTACCAACGACCCGCGACGCGCAAGAACTTCACTCTCCACTGTCAAGCTTCC	1500
I S Y Q R C T A D S A R T F I S T V K L P	
CGAAACCAAGAAGTGAATCTAAACCCCTCTCCACCGTCAAACCTGCAACTGTTGCTCT	1560
E T T S E S K T P P P T V K P A T V A L	
TCCTTATCGTTACATTCCCGGTAAACAAAACAACTTAACCCAAATCCCCGACCGACTAGTGA	1620
P Y R Y I P G N K T T -	
TTAATTTTACTAGCAATTAGTAATTTGTAATTAGGCAATCAAAATGGCAATACCTCTTC	1680
M A I P L	
TGCAATACAAACCTAGTTCCCAAAATCAACGCGTTCGCGGTTACGAAGTACCCAAACGAAG	1740
L Q Y K P S S Q N Q R V P G Y E V P N E	
ATACCCCGCAAAATTTATCGCTTAGAAGATGCTCCGAGTTATAGTGAATTAAGAAGTAA	1800
D T P R I Y R L E D A P T S Y S E I Q E L	
TTTGGCGAGCTTACCGACAAATATTGACGGAACACGAAATCCTCAAATTCACCCGCCAAA	1860
I W A A Y R Q I F S E H E I L K F H R L	
TCAATCTAGA 1870	
I N L	

		10	20	30	40
$L_R^{31.5}$	(<i>pec</i> C)	M-STSV A ERL AIKD-EVD K I ELRPNWSED E LQIVFKTAY			
$L_R^{32.3}$	(<i>cpc</i> C)	M AITAAASRL GTEPF S NA A K I ELRSDASRE E VEAVINAVY			
					• • *
		50	60	70	80
$L_R^{31.5}$	(<i>pec</i> C)	EQVFGRQGLY A SQRFATA E A L LRNGKISVK Q FI E LLAKSE			
$L_R^{32.3}$	(<i>cpc</i> C)	RHVL G NDYIM A SERL V SA E S L LRDGNLT V R E FRSV A KSE			
		• •	*	• • • • •	• • • • •
		90	100	110	120
$L_R^{31.5}$	(<i>pec</i> C)	FYKE C FFYNN S -QVR F IELN YK H LLGRAPY DQSE I AFHVD			
$L_R^{32.3}$	(<i>cpc</i> C)	LYKK K FFY-N S FQ T R F IELN YK H LLGRAPY DESE I VFHLD			
		•	• •	• • • • •	*
		130	140	150	160
$L_R^{31.5}$	(<i>pec</i> C)	LYAAAGY D AE I ESY I YSPEY D N AFGNF V VP Y YRG F QS I PG			
$L_R^{32.3}$	(<i>cpc</i> C)	LYQN K GY D AE I DSY I DSVEY Q N NFGDN I VP Y YRG F ETQ P G			
		*	• • • • •	**	** • *
		170	180	190	200
$L_R^{31.5}$	(<i>pec</i> C)	MKT V GFNR I F E LYRGRANSD N AQFGGKSAR L RSKISM N LA			
$L_R^{32.3}$	(<i>cpc</i> C)	QKT V GFNR M F R LYRGYANSD R AQ I EGTKPR L ARELAT N KA			
		210	220	230	240
$L_R^{31.5}$	(<i>pec</i> C)	NTIVPPT S PI A A----- S T S SAR- T LVTS P V M -GDAR M F			
$L_R^{32.3}$	(<i>cpc</i> C)	S SIVGP S GSN P AWGYR P SVD I TPRK T LG N - A V G -ENDRVY			
$L_R^{8.2}$	(<i>cpc</i> D)	M FGQT T LG I D S VSSASRVF			
$L_C^{7.7}$	(<i>apc</i> C)	M GR L F			
		250	260	270	280
$L_R^{31.5}$	(<i>pec</i> C)	IVEAIAGTLN T-----N V A V RR S RQVY-T V PYDRLSATY			
$L_R^{32.3}$	(<i>cpc</i> C)	RIEVTGVRSP G-----Y P S V RRSSYA-II V PYERLSEKI			
$L_R^{8.2}$	(<i>cpc</i> D)	RFEVVGMRQN EENDKNKY-N I RRSGSVYIT V PYNRMSEEM			
$L_C^{7.7}$	(<i>apc</i> C)	KITACVPSQT R----IRTQ R E LQSTYFTKL V PYENWFREQ			
		290	300	310	
$L_R^{31.5}$	(<i>pec</i> C)	Q E I HKRGG K I VKITPAS			
$L_R^{32.3}$	(<i>cpc</i> C)	Q Q I HKLGG K I VSITSA			
$L_R^{8.2}$	(<i>cpc</i> D)	Q R I HRLGG K I VKIEPL-TRA AG			
$L_C^{7.7}$	(<i>apc</i> C)	Q R I Q M GG K I VKVE-LATGK QGINTGLA			

		10	20	30	40
Repeat A	(<i>apc</i> E)	-PSNKVRQR-	PSGDQQGLE-	LPQIYFNAAE	RR-QKFVMKP
Repeat B	(<i>apc</i> E)	NP-N-ARGEF	PGSLGAKVFR	LDQIPRTIS-	KK-TGKGASI
Repeat C	(<i>apc</i> E)	GAVTETRTE-	P--DVQ--FR	INQGVSKKRE	QR-KIFKLVA
Repeat D	(<i>apc</i> E)		GRGQSVE-	VGVTTRRK	ARIYRYRMTV
		50	60	70	80
Repeat A	(<i>apc</i> E)	GLSA-AEKNE	VVKAAYRQIF	E--RDITRAY	SLSISY--LE
Repeat B	(<i>apc</i> E)	KFSE-SSTQA	VIKAAYLQVF	G--RDVYEQ	RLKVQ---E
Repeat C	(<i>apc</i> E)	GTTDKVAVQN	TISAAYRQIF	E--RDIAPYV	ASSREFKVLE
Repeat D	(<i>apc</i> E)	GTNQ-AEIQK	IIDAIIYIQVM	DVFSGQVPEY	-FRRPD--LE
		90	100	110	120
Repeat A	(<i>apc</i> E)	SQVKNQDISM	KEFVRRRLGKS	PLYRKQFYEP	FINSRALELA
Repeat B	(<i>apc</i> E)	IKLENQEISV	REFVRALAKS	DLFRKLYWTP	LYVCKAIEYI
Repeat C	(<i>apc</i> E)	SKLSNQEITV	KEFIEGLGCS	GLYLKEFYAP	YPNTKVIELG
Repeat D	(<i>apc</i> E)	SKLRNQEISV	REFVRDLASS	EIYRKRFYTP	YPNTKVIIEFL
		130	140	150	160
Repeat A	(<i>apc</i> E)	FRHFLGRGPS	SREEVQKYFD	IISKGGLPAL	VDALVDSQEY
Repeat B	(<i>apc</i> E)	HRRLGRPTY	GRQENNKYFD	ICAKKGFYAF	IDAIIDSEY
Repeat C	(<i>apc</i> E)	TKHFLGRAPL	DQAEIRKYNQ	ILATOGIRAF	IRAMLNTQY
Repeat D	(<i>apc</i> E)	FRHLLGRAPA	TQAEIRQYNK	LLADGGLKAA	VEAMVNSPEY
		170	180	190	200
Repeat A	(<i>apc</i> E)	SDYFGEETVP	YIRGLGQEAQ	ECRNWGPQQ-	DLFKYSAPFR
Repeat B	(<i>apc</i> E)	SQAFGEDTVP	YERYLTPAGV	ALRKL RVGS-	IREDEVKAKVE
Repeat C	(<i>apc</i> E)	AEAFGEDTVP	YNRFPTLP		
Repeat D	(<i>apc</i> E)	TRYFGEDVVP	YPRFPSLPAG	NYLGSVKAAA	DLVKQSWSSL
		210	220	230	240
Repeat A	(<i>apc</i> E)	KV--PQFIT	FADYERPL-P	DQHPYSGGND	PL
Repeat B	(<i>apc</i> E)	KEEKPRFVEL	GAVTETRTEP	DVQFRINQGV	SKKREQRKIF
Repeat C	(<i>apc</i> E)				
Repeat D	(<i>apc</i> E)	S---PSVLTG	RYSQS		

Figure 2. Sequence comparison of the class I and class II linkers of the phycobilisome from *Mastigocladus laminosus* (Fischerella PCC7603), and of the four repeats of the core-membrane linker L_{cm} . Those parts used for the homology analysis shown in Table 1 are marked by different shading: domain 1 is indicated by a light gray background, domain 2 is boxed. The ... and *** indicate overall homologies and identities, respectively, between the N-terminal regions (or repeats) of all the linkers. However, for clarity of display those linkers showing high degrees of homology are grouped.

A: rod class II linkers and class I linkers, B: rod-core linkers, and C: linker-like repeats (domain-type 1) A to D of the L_{cm} .

(a) The domain-boarders found differ slightly from those reported. The “repeats” of the $L_{cm}^{127.6}$ give higher homologies if they are extended (A, B, and C are extended by 25 to 50 amino acids in N-terminal direction, repeat C by 20 amino acids in the C-terminal direction). These extensions lead to shorter “arm” regions between the repeats, but include the highly conserved core regions.

(b) A homology between the small linkers (especially $L_c^{7.7}$) and the C-terminal part of the $L_{cm}^{127.6}$ (denoted as “arm 5” in Esteban, 1993) could not be detected. This might be due to the different domain boarders used for the comparative sequence analysis, or by differences in stringency for the alignment programs. C-terminal extension of the domain boarder is supported by isolation of an APC- L_{cm} fragment complex corresponding to repeat D (Gottschalk *et al.* 1993), but extending at least to aa 1116, which is 22 aa longer than the boarder given in Esteban 1993.

This defines what we call domain 1 which comprises the N-terminal ~ 22 kDa portion of the class II linkers, viz. of the rod ($L_r^{32.3}$ and $L_r^{31.5}$) and rod-core linkers ($L_{rc}^{31.8}$, $L_{rc}^{28.7}$, and $L_{rc}^{29.6}$), and the repeats in the L_{cm} . The homology (based on identity of amino acids) in this domain is 30.9 to 69.1%. It is even higher within the two rod-linker sub groups, *e.g.*, among the three L_{rc} (59.1 to 70.2%) and among the two L_r (54.3%) and amounts to 39 to 45% among the L_{cm} repeats.

The domain boarders found in this analysis can be compared with linker fragments in the 20 kDa range, obtained by proteolysis of biliprotein-linker complexes. An N-terminal 22 kDa fragment of $L_{rc}^{31.8}$ corresponding to this domain 1 has been identified in a PC-trimer with a strongly red-shifted absorption spectrum (Gottschalk *et al.* 1991). Its C-terminus has not been determined, but is expected to from the mobility in SDS-PAGE about 10 aa C-terminal from domain boarder. This value is based on an apparent size of 115 ± 1 Da/aa in the three L_{rc} proteins. More recently, the fourth, C-terminal repeat of $L_{cm}^{127.6}$ (22 kDa) has been isolated as a component of an APC-trimer complex (Esteban *et al.* 1990, Esteban 1993, Gottschalk *et al.* 1993). This fragment extends at least to aa 1116 (from C-terminal protein sequence analysis, Gottschalk *et al.* 1993, see above). In both cases the portion of the linker isolated with the biliprotein corresponds to domain 1, which is most likely protected from proteolysis because it is buried in the central hole of the doughnut-shaped trimer.

The homology among the five class II linkers sharply decreases in the remaining C-terminal parts of the molecules. However, the two rod linkers are homologous among themselves with this region forming a second additional domain 2. This domain 2 of the rod linkers has considerable homology to the class I linkers, viz. to $L_r^{8.2}$ (40.3 to 40.9%) and even to the related $L_c^{7.7}$ APC linker (28.3 to 32.1%). Since these two small linkers are also homologous to each other, they are grouped in domain 2 which is then present in all rod linkers (L_r) irrespective of their internal (L_r) or distal-terminal po-

sition ($L_r^{8.2}$). We use the functional designation of Lundell *et al.* (1981), whose experiments have shown that increasing proportion of $L_r^{8.2}$ inhibits the elongation of rods, and have therefore termed it rod-terminating linker. The homology of $L_c^{7.7}$ suggests that the latter may perform a similar capping function in the short rods of which the core is build up.

The volume of the central hole of a hexameric disk of any of the biliproteins has been calculated to be in the range of one monomer (~34 kDa) (Ficner *et al.* 1992, Ficner and Huber 1993). Any of the class II rod or rod-core linkers could therefore fit well into a single such hexamer. However, the presence of domain 2 in the C-terminal part of the rod-linkers and (over almost the entire length) of the capping linkers $L_c^{7.7}$ and $L_r^{8.2}$ suggests a different arrangement of the rod components which is shown in Figure 3. Starting at the distal end of a rod, the capping linker ($L_r^{8.2}$) fills one third of the distal PC (or PEC) hexamer's central hole, leaving space for only ~22 kDa of the next linker. This space is taken up by the N-terminal 22 kDa part (~domain 1) of the outermost rod linker (L_r), which therefore protrudes from the hexamer with its C-terminal domain 2. This in turn fits (as the homologous $L_r^{8.2}$) into the second hexamer, leaving again space for the N-terminal domain 1 of the next rod linker.

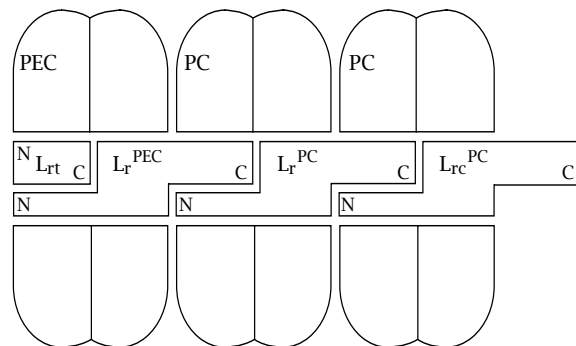


Figure 3. Schematic representation of the interlocking rod model, for phycobilisome rods from *M. lamosus*, shown for a rod containing one PEC and two PC hexameric disks. The linkers are generalized as L_{rt} , L_r^{PEC} , L_r^{PC} and L_{rc}^{PC} corresponding to $L_r^{8.2}$, $L_r^{31.5}$, $L_r^{32.3}$, and the three rod-core linkers $L_{rc}^{31.7}$, $L_{rc}^{28.7}$, or $L_{rc}^{29.6}$, respectively. The orientation is indicated by “N” for the N- and “C” for the C-terminus. The C-terminal protrusion of L_{rc} at the right site, is responsible for attachment to the core.

The asymmetric arrangement of the small rod-terminating linker (L_{rt}) has been taken in analogy to the respective APC-linker; this is propagated along the rod by the terminal portions of the rod- and rod-core linkers.

This is repeated in the third hexamer (assuming a three-disk rod), but now the rod-core linkers fit with their N-terminal (and homologous) domains 1 into the remaining space. The C-terminal portions of these link-

ers differ considerably among each other and from the $L_r^{32.3}$ and $L_r^{31.5}$. This provides the structures for attaching the rods to the APC core, where the orientation of the adjacent disks is perpendicular rather than parallel as in the rods, and also provides the specificity of the docking positions.

It should be noted that the size of the suggested domain 1 fits the space of one single trimer only incompletely (space in the interior hole of $\alpha_3\beta_3$ is 16 to 18 kDa). The first part of the N-terminal domain 1 is then perhaps responsible for building up the stable hexamer ($\alpha_6\beta_6$)-linker complex in some cyanobacteria (leaving out strains where stable hexamers without linkers could be detected). This function of the first part of domain 1 may account for the reduced homology within the first 40 amino acids of the rod and rod-core linker polypeptides reported by Glauser (1991). This is supported by the X-ray structure of an AP-linker complex ($\alpha_3\beta_3 L_c^{7.7}$). The linker-polypeptide corresponding to a type 2-domain, fills only 2/3 of the central cavity of the phycobiliprotein trimer, leaving space for parts of other linker polypeptides (Reuter *et al.* 1999).

A domain structure could also be discerned in the chromophore-containing γ -subunit of a marine cyanobacterium. Wilbanks and Glazer (1994) suggest a two (or three) domain structure, consisting of:

- (a) an N-terminal 50 amino-acid chromophore containing domain (in size nearly identical to the small $L_r^{8.2}$ and $L_c^{7.7}$ and domain 2 of the rod and rod-terminating linkers, but without significant sequence identity), and of
- (b) a linker domain 1 type stretch of ~200 amino acids, showing a homology to the phycocyanin and phycoerythrin associated linker polypeptides. A complete domain 2, as represented in the rod linkers and the small linkers, is missing, and only the last 25 amino acids of the C-terminus could be interpreted as the beginning of domain 2 (the facile proteolytic cleavage of this part underlining this suggestion). These results fit very well with the observed X-ray structure from Ficner and Huber (1993) where no electron density could be detected outside of the hexameric disk.

The proposed model for the rods might be extended to the core as well, with the repeating domains 1 of $L_{cm}^{127.6}$ filling the interior parts of the APC holes, and $L_c^{7.7}$ (domain 2) capping them.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, Bonn (SFB 143 and 5FB 533, Teilprojekt A1).

REFERENCES

- [1] F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, *Current Protocols in Molecular Biology*, John Wiley and Sons, New York, 1987.
- [2] D. A. Bryant, G. Guglielmi, and N. Tandeau de Marsac, *Arch. Microbiol.* **123** (1979), 113.
- [3] R. W. Castenholz, *Schweizer Z. Hydrol.* **35** (1970), 538.
- [4] M. Düring, R. Huber, W. Bode, R. Rübli, and H. Zuber, *J. Mol. Biol.* **211** (1990), 633.
- [5] M. Eberlein and W. Kufer, *Gene* **94** (1990), 133.
- [6] A. M. Esteban, Ph.D. thesis, ETH Zürich, 1993.
- [7] A. M. Esteban, V. Stirewalt, G. Frank, D. A. Bryant, and H. Zuber, *First European Workshop on the Molecular Biology of Cyanobacteria*, Dourdan, 1990, Book of Abstracts (N. Tandeau de Marsac *et al.* eds), pp. 41.
- [8] C. D. Fairchild and A. N. Glazer, *J. Biol. Chem.* **269** (1994), 8686.
- [9] R. Ficner and R. Huber, *Eur. J. Biochem.* **218** (1993), 103.
- [10] R. Ficner, K. Lobeck, G. Schmidt, and R. Huber, *J. Mol. Biol.* **228** (1992), 935.
- [11] R. Fischer, J. Gottstein, S. Siebzehrnühl, and H. Scheer, *Perspectives in Photosynthesis*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1990, (J. Jortner and B. Pullman eds), pp. 121-131.
- [12] A. M. Frischauf, H. Lehrach, A. Poustka, and N. Murray, *J. Mol. Biol.* **170** (1983), 827.
- [13] E. Gantt, *Photosynthesis III*, Springer Verlag, Berlin, 1986, (L.A. Staehelin and C. J. Arntzen eds), pp. 260-268.
- [14] M. Glauser, Ph.D. thesis, ETH Zürich, 1991.
- [15] M. Glauser, V. L. Stirewalt, D. A. Bryant, W. Sidler, and H. Zuber, *Eur. J. Biochem.* **205** (1992), 927.
- [16] A. N. Glazer, *Biochem. Biophys. Acta* **768** (1984), 29.
- [17] L. Gottschalk, R. Fischer, F. Lottspeich, and H. Scheer, *Photochem. Photobiol.* **54** (1991), 283.
- [18] L. Gottschalk, F. Lottspeich, and H. Scheer, *Photochem. Photobiol.* **58** (1993), 761.
- [19] D. Hanahan, *DNA cloning, a practical approach*, Oxford IRL, Oxford, 1985, (Glover, D.M. ed), pp. 109-135.
- [20] P. Heinrich, in *Guidelines for Quick and Simple Plasmid Sequencing*, Boehringer Mannheim GmbH, Mannheim, 1986.
- [21] D. G. Higgins and P. M. Sharp, *Gene* **73** (1988), 237.
- [22] R. H. Lau, G. Alvarado-Urbina, and P. C. K. Lau, *Gene* **52** (1987), 21.
- [23] D. J. Lundell, R. C. Williams, and A. N. Glazer, *J. Mol. Biol.* **256** (1981), 3580.
- [24] A. Manadori and A. Melis, *FEBS Lett.* **181** (1985), 79.
- [25] T. Maniatis, E. F. Fritsch, and J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1982.
- [26] D. Mazel, G. Guglielmi, J. Houmard, W. Sidler, D. A. Bryant, and N. Tandeau de Marsac, *Nucl. Acid. Res.* **14** (1986), 8279.
- [27] E. Mörschel, K. P. Koller, W. Wehrmeyer, and H. Schneider, *Eur. J. Cell Biol.* **16** (1977), 118.

-
- [28] W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. USA* **85** (1988), 2444.
- [29] G. Porter, C. Tredwell, G. F. Searle, and J. Barber, *Biochem. Biophys. Acta* **501** (1978), 232.
- [30] W. Reuter and C. Nickel-Reuter, *J. Photochem. Photobiol. B: Biol.* **18** (1993), 51.
- [31] W. Reuter, G. Wiegand, R. Huber, and M. Than, *Proc. Natl. Acad. Sci. USA* **96** (1999), 1363.
- [32] H. Scheer, *Angew. Chem.* **93** (1981), 230, *Angew. Chem. Int. Ed.* **20**, 241.
- [33] T. Schirmer, W. Bode, and R. Huber, *J. Mol. Biol.* **196** (1987), 677.
- [34] T. Schirmer, R. Huber, M. Schneider, W. Bode, M. Miller, and M. Hackert, *J. Mol. Biol.* **188** (1986), 651.
- [35] W. Sidler, J. Gysi, E. Isker, and H. Zuber, *Hoppe-Seyler's Z. Physiol. Chem.* **362** (1981), 611.
- [36] N. Tandeau de Marsac, D. Mazel, T. Demerval, G. Guglielmi, V. Capuano, and D. J. Houmard, *Photosyn. Res.* **18** (1988), 99.
- [37] S. M. Wilbanks and A. N. Glazer, *J. Biol. Chem.* **268** (1993), 1236.
- [38] C. Yanish-Perron, J. Vieira, and J. Messing, *Proc. Natl. Acad. Sci. USA* **80** (1985), 1194.
- [39] H. Zuber, *TIBS* **11** (1986), 414.
- [40] K. H. Zhao, M. G. Deng, M. Zheng, M. Zhou, A. Parbel, M. Storf, M. Meyer, B. Strohmman, and H. Scheer, *FEBS Lett.* **489** (2000), 9.

