Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*

W.R.Belknap and R.Haselkorn

Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA

Communicated by G.Tocchini-Valentini

The biliprotein phycocyanin (PC) is a major constituent of the light-harvesting apparatus of cyanobacteria and red algae. A DNA fragment encoding the β and α subunits of PC was isolated from a genomic library of the cyanobacterium *Anabaena* 7120 DNA. The single-copy PC genes are part of a larger operon which consists of five open reading frames (ORFs) encoding, in order, the β and α subunits of PC, two linker polypeptides associated with PC in phycobilisome rods, and a fifth ORF, which may encode a linker polypeptide involved in attachment of the phycobilisome rod to the core of the structure. The operon yields three major transcripts, the first of which (1.4 kb) encodes only the PC subunits. A second (3.6 kb) encodes all five ORFs, and appears to arise from partial read-through of a terminator following the PC subunit genes. The third transcript (1.4 kb) encodes the last two ORFs. The relative levels of the three transcripts *in vivo* are modulated by light intensity, but they are not altered by the removal of fixed nitrogen from the growth medium. The site of light regulation appears to be the terminator following the PC genes, rather than a promoter.

Key words: phycobiliprotein/linker proteins/DNA sequence/transcription/RNA secondary structure

Introduction

Phycobilisomes are large protein complexes which serve as the light-harvesting antennae for photosystem (PS) II in cyanobacteria and red algae (Glazer, 1985; Gantt, 1981). These highly ordered structures are composed of two protein types: the pigmented biliproteins which are involved directly in light absorption and energy transfer, and non-pigmented linker polypeptides which serve structural roles (Tandeau de Marsac and Cohen-Bazire, 1977). The family of biliproteins is made up of members characterized by different absorption and fluorescence maxima. Each of these members is composed of two non-identical polypeptide chains (α and β) which contain covalently linked open-chain tetrapyrroles as chromophores. Phycobilisomes are constructed of rods composed of stacked discs radiating from central cores which are associated with the thylakoid membranes (Bryant *et al.*, 1979). The basic unit of phycobilisome assembly is a hexameric α - β double disc with an associated linker polypeptide, $[(\alpha\beta)_3]_2\gamma$. The biliproteins are organized within the phycobilisome such that members with short wavelength absorption maxima [phycoerythrin (PE) or phycoerythrocyanin (PEC)] are located at the terminal regions of the rods, biliproteins with intermediate maxima [phycocyanin (PC)] are located in the rods proximal to a core containing allophycocyanin (APC), which absorbs at a still longer wavelength. The phycobilisome is anchored to the thylakoid membrane by a large biliprotein which serves

as the terminal acceptor for the structure, and transfers the light energy to PS II (Lundell *et al.*, 1981). This organization permits very efficient (>90%) energy transfer from peripheral PE or PEC components to the PS II reaction centers (Porter *et al.*, 1978; Searle *et al.*, 1978). The phycobilisome structure is stabilized by a variety of linker polypeptides, each of which is associated with a specific biliprotein and has a defined location within the complex.

We have isolated the genes coding for the phycocyanin operon from the filamentous cyanobacterium *Anabaena* 7120. This cyanobacterium differentiates cells specialized for dinitrogen fixation, called heterocysts, in response to deprivation of a fixed nitrogen source (Haselkorn, 1978). In addition to heterocyst induction, the removal of fixed nitrogen also results in yellowing of the cultures, which reflects a decreased phycobiliprotein content in the filaments. The induction process has been shown to result in an increase in phycobiliprotein-specific protease activity both in vegetative cells and developing heterocysts (Wood and Haselkorn, 1980). As the phycobilisomes in cyanobacteria constitute a substantial fraction of the total soluble protein (up to 60%), the catabolism of these structures would supply reduced nitrogen for the filaments in the period prior to the formation of mature, nitrogen-fixing, heterocysts.

In addition to changes in biliproteins in response to nutrient availability, the cyanobacteria and eukaryotic algae can alter both the number and relative composition of phycobilisomes in response to changes in incident light intensity (quantity) (Kratz and Myers, 1955; Oquist, 1974a,b) and wavelength (quality) (Bogorad, 1975). The regulation of cyanobacterial phycobilisome composition by light quality, referred to as chromatic adaptation, has been most thoroughly investigated in *Fremyella diplosiphon* (Grossman *et al.*, 1986). This regulation has been demonstrated to involve differential expression of the component biliprotein genes (Conley *et al.*, 1985, 1986). In contrast, transcriptional regulation of biliprotein genes by light quantity has not been conclusively shown (Gasparich, *et al.*, 1987).

The genes coding for biliproteins and linker polypeptides have been isolated from a variety of cyanobacteria and eukaryotic algae. In all cases, DNA coding sequences for the α and β subunits of the biliproteins are closely linked. When the transcripts originating from these sequences have been analyzed, co-transcription of the α and β subunits is observed (Pilot and Fox, 1984; Conley *et al.*, 1985, 1986; Lemaux and Grossman, 1985; Houmard *et al.*, 1987). This transcriptional property could provide the required 1:1 ratio of the α and β subunits within the phycobilisome. Longer transcripts containing sequences coding for linker polypeptides in addition to biliproteins have also been observed (A.Grossman and T.Lomax, personal communication).

The *Anabaena* PC operon consists of five open reading frames (ORFs) which encode the PC subunits as well as linker polypeptides, with a structure similar to that found in *Synechococcus* 7002 (Bryant *et al.*, 1987). We have identified three transcripts originating from the operon. Starvation for fixed nitrogen did

not alter the transcription of the genes in this operon. Transcription of the PC operon is, however, regulated by light intensity. This regulation appears to involve attenuation of transcriptional termination or post-transcriptional processing, rather than light-regulated transcriptional initiation.

Results

Isolation of the *Anabaena* 7120 PC genes

The biliproteins are highly conserved among cyanobacteria and eukaryotic algae. Genes isolated from one species have previously been used as heterologous hybridization probes to isolate genes from other organisms (Conley *et al.*, 1985; Houmard *et al.*, 1987). We used DNA encoding PC from the eukaryotic alga *Cyanophora paradoxa* cyanelle as a probe to isolate PC genes from *Anabaena*.

A probe consisting of a 0.8-kbp *Pst*I–*Eco*RI fragment from pCPC3029 (Lemaux and Grossman, 1984, 1985), which contains part of the coding regions for both the α and β subunits of *C. paradoxa* PC, was used to probe a Southern blot of total *Anabaena* 7120 DNA under conditions of low stringency (Figure 1a). Single prominent bands were observed in three of the restriction digests, while two bands were observed in the *Xba*I digest. Later sequence analysis of the PC genes of *Anabaena* 7120 revealed the presence of both *Hind*III and *Xba*I restriction sites within the coding sequence of PC β . Other bands were observed to hybridize with the probe in each digest with considerably less intensity. These low intensity bands may reflect hybridization of the probe to *Anabaena* PEC or APC genes (Lemaux and Grossman, 1985).

The *C. paradoxa* probe was then used to screen a cosmid library of *Anabaena* 7120 DNA. A positive clone (pAn400) with an insert of ~ 35 kb was obtained and used in further analysis. Southern blots of pAn400 DNA probed with *Cyanophora* PC genes yielded a hybridization pattern similar to that of Figure 1 (data not shown). Additional DNA bands hybridizing to a lesser

Fig. 1. Hybridization of PC genes from *C. paradoxa* (a) and *Anabaena* 7120 (b) to total *Anabaena* 7120 DNA. *Anabaena* DNA ($\sim 3 \mu\text{g}$) was digested with *Hinc*II (1), *Hind*III (2), *Xba*I (3) or *Eco*RI (4), electrophoresed on a 0.8% gel and transferred to a nylon membrane. (a) DNA hybridized at low stringency with a nick-translated 0.8-kbp *Pst*I–*Eco*RI fragment of pCPC3029 (Lemaux and Grossman, 1984) containing parts of both the PC α and β coding sequences of *C. paradoxa*. (b) DNA hybridized at high stringency to a nick-translated 190-bp *Hinc*II–*Xba*I fragment of pAN420 (Figure 4) DNA containing part of the *Anabaena* 7120 PC β subunit coding region.

degree were observed, which may indicate linkage of the PC coding region to that of other biliproteins.

A 230-bp *Hinc*II–*Hind*III fragment containing coding sequences of the β subunit of *Anabaena* PC was prepared, and used to probe total *Anabaena* DNA (Figure 1b). A hybridization pattern identical to that obtained with the heterologous probe was obtained. These data indicate that the PC gene is single copy in *Anabaena* 7120.

Nucleotide sequence of *Anabaena* 7120 PC genes and downstream regions

Hybridization of an *Anabaena* PC gene probe to blots of total *Anabaena* 7120 RNA prepared from both ammonium-replete and ammonium-starved cultures (Figure 2) indicated the presence of a large (3.6 kb) message in addition to a shorter (1.4 kb) hybridizing RNA. A similar hybridization pattern has been observed in the expression of the red light-induced PC genes from *F. diplosiphon* (Conley *et al.*, 1985, 1986). In *Fremyella*, the longer transcript has been shown to encode phycobilisome linker protein sequences downstream from the PC genes (A. Grossman and T. Lomax, personal communication). A similar gene arrangement has been found for the PC genes in *Synechococcus* 7002 (Bryant

Fig. 2. Transcript analysis of the *Anabaena* PC operon. Northern blots of total *Anabaena* RNA prepared from cultures uninduced (U) and induced to differentiate heterocysts (I) were hybridized with (a) a 190-bp *Hinc*II–*Xba*I fragment internal to the β subunit of PC, (b) a 140-bp *Rsa*I–*Hind*III fragment internal to the *cpcC* gene and (c) a 850-bp *Xba*I–*Hind*III fragment containing the 3' end of the *cpcE* gene. The location of the DNA probes relative to the PC operon ORFs is shown at the top of the figure. Transcript sizes were determined using RNA markers (Bethesda Research Laboratories).

et al., 1987). To determine the corresponding structure in *Anabaena*, both the PC hybridizing regions and downstream DNA were sequenced. The nucleotide sequence of the operon and deduced amino acid sequences of the ORFs are shown in Figure 3. Two subclones of pAn400 were utilized for this analysis. The plasmid pAn410 (Figure 4) contains the 4.3-kbp *HincII* PC hybridizing band (Figure 1). It encodes part of the β subunit of PC, all of the α subunit and the downstream regions co-transcribed with *Anabaena* PC. The plasmid pAn420 contains a 1.5-kbp *HindIII*–*XbaI* fragment of *Anabaena* DNA encoding the amino terminus of the β subunit of PC and the upstream flanking sequences. The sequencing strategy used is indicated in Figure 4.

Five ORFs with appropriate ribosome binding sites were found within the sequence. Four of the ORFs could be unequivocally assigned to proteins by comparison with published protein and DNA sequences.

The β and α subunits of PC are encoded in the first two ORFs of the operon (*cpcB* and *cpcA*, Figure 4). The subunits are in the same orientation (β – α) observed with other cyanobacterial PC genes (Pilot and Fox, 1984; de Lorimier *et al.*, 1984). Identical possible ribosome binding sites (AGGAGA) are found upstream from both coding regions (9 and 10 bp upstream for β and α respectively) (Figure 3). There is a 92-bp untranslated region between the β and α coding sequences. In Figure 5 the deduced amino acid sequences of the *Anabaena* 7120 PC subunits are compared to those of three other cyanobacteria. A high degree of homology (~55%) is observed among the PC genes. Regions of complete homology include sequences surrounding the cysteinyl residues involved in bilin linkage (α subunit residue 85, β residues 83 and 154).

The *cpcA* gene is followed by an inverted repeat, the transcription of which will result in RNA capable of folding into the secondary structure shown in Figure 9a. This proposed secondary structure has a calculated ΔG of ~ -43 kcal/mol (Salser, 1977). The structure begins 9 bp downstream from the stop codon for the PC α subunit, but is not followed by the string of T's that usually characterize a transcriptional terminator (Platt, 1986).

The third ORF in the sequence, *cpcC*, begins 140 bp downstream from the last codon for the α subunit of PC (Figure 4). It is preceded by a potential ribosome binding site (AGG-GGA) 8 bp upstream from the first codon. Based on comparison with previously published sequence (Figure 6), this ORF (*cpcC*) can be assigned to the 32-kd rod linker associated with PC, $L_{R}^{32,PC}$ in the notation of Glazer (1985). The rod linker is the protein that stabilizes PC hexameric discs $[(\alpha\beta)_3]_2$ in the phycobilisome rod. The first 46 residues of the PC rod linker of *Mastigocladus laminosus* (Ma) have been determined (Fuglistaller *et al.*, 1985), and a good fit to the deduced N-terminal amino acid sequence of the *cpcC* ORF is observed (Figure 6). The amino acid sequences of both the amino and carboxy termini of the rod linker associated with PEC in *Mastigocladus* have also been determined (Fuglistaller *et al.*, 1985). Regions of high homology to the *cpcC* ORF are observed both in the center of the ORF (residues 161–207) and the carboxy terminus of the linker (residues 252–287).

The fourth ORF, *cpcD*, (Figure 7) begins 134 bp downstream from the end of the *cpcC* gene. It is preceded by a potential ribosome binding site (AGGAGA, 10 bp upstream from the start codon) and can be assigned to the 8.9-kd rod-capping polypeptide ($L_{R}^{8.9,PC}$) (Glazer, 1985). Excellent homology is observed between the deduced amino acid sequence of this ORF and the sequence of the rod-capping protein from *M. laminosus*

(Fuglistaller *et al.*, 1986a,b) (Figure 7).

The deduced amino acid sequence of the fifth ORF, designated *cpcE*, is presented in Figure 8. This ORF begins 15 bp downstream from the end of the *cpcD* gene. The polypeptide synthesized by translation of the ORF starting from the initial methionine would have a mass of 30 kd. However, this methionine is not preceded by a ribosome binding site (the sequence of the 20 bp upstream is GGCATAATAGCGCCCTG-GTA). On the other hand, 150 bp downstream from the end of the *cpcD* gene, the *cpcE* ORF contains a valine coded for by GTG (residue 46, Figure 8). This valine is preceded by a potential ribosome binding site (AGGTAA, 7 bp upstream). If translation of this ORF begins at this valine residue, the resulting polypeptide would have a molecular mass of ~ 25 kd.

The deduced amino acid sequence of *cpcE* does not have a high degree of homology with any published phycobilisome linker sequence, but it does have properties generally associated with phycobilisome linkers. It has been observed that linker sequences have homologies to the subunits of the biliproteins (Fuglistaller *et al.*, 1985). In Figure 8, the deduced amino acid sequence of *cpcE* is shown aligned with sequences from the subunits of *Anabaena* 7120 PC. In this case it can be seen that the putative amino terminus of the ORF bears homology to the α subunit of PC. Homology to PC β sequences is also observed in the ORF (residues 76–95), as well as other sequences homologous to the α subunit. These homologies suggest that the *cpcE* ORF codes for a linker associated with PC.

In *Anabaena variabilis* phycobilisomes, PC is associated with two large linkers (27 and 32 kd) and a smaller (9 kd) linker (Yu *et al.*, 1981; Yu and Glazer, 1982). The 32-kd linker stabilizes PC subunits within the phycobilisome rods, while the 27-kd polypeptide ($L_{RC}^{27,PC}$) (Glazer, 1985) links the phycobilisome rod to the core. We can assign ORFs *cpcC* and *cpcD* by sequence homology to the rod and cap linkers associated with PC. Thus it is reasonable to assign the *cpcE* coding region with the remaining linker associated with PC, $L_{RC}^{27,PC}$. The assignment of the final ORF to the rod–core linker is therefore based on its size, its sequence homology to the biliprotein subunits, and its location in the operon.

The *cpcE* ORF is followed immediately by an inverted repeat, the transcription of which will result in RNA capable of forming the secondary structure shown in Figure 9b. The calculated ΔG of this structure is -37 kcal/mol (Salser, 1977). Although this structure is not followed by a string of T's in the direction of transcription, we believe that it represents the termination signal for the 3.6-kb transcript.

Transcription of the *Anabaena* 7120 PC operon

Growth of *Anabaena* 7120 in the absence of a fixed nitrogen source results in the induction of heterocysts (Haselkorn, 1978). These cells, specialized for the fixation of atmospheric dinitrogen, occur at regular intervals along the cyanobacterial filaments. The differentiation process involves not only transcriptional activation of genes required for nitrogen fixation, but also the loss of several photosynthetic functions associated with vegetative cells. These lost functions include PS II and ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activities. It has been suggested that even though PS II functions are lost in mature heterocysts, these cells retain PC (Yamanaka and Glazer, 1983), which may function in energy transfer to photosystem I (PS I) (Peterson *et al.*, 1981). To analyze possible changes in transcription of PC and linker genes during heterocyst differentiation, RNA was prepared from both induced (growth for 36 h in the absence of

-479 AAAAAATGGCGTAAAAACAGCAATAGATTGTCACCACTAGGTTTTTGCAACTATTAAAAATAATTAAGACACCAATATTT
 -399 TGCCTAAAATCACTAATTCATGAAGTTATGTTACGGGAGAACGGCGTTTTTCTCAGTTCTAGCCCTTGATGTAAGCAT
 -319 TTGAGGGGGTTCTTATTCATGCAACTTATTTATTACAAATTTGTAACAAAATAAGGATCTATAGCATTTGTATAAACATAA
 -239 GCTGGAGGGGTTAAACGACAGACAAAAGTTAAACGAGGCAATCGATTACAAAAGTTTGCCTGCTTCTCATAAAATCACCC
 -159 CAGTTTCTCATTTTTTCTCATTTCATGGAAGTGGCAAGAAATTTTGAGACTAACCTGAGAAGTAGTTTGATCCCTGGTCTGT
 -79 TTACAACCGTTTCCAAGAAATAAACATAGCCAGTTTAAATTCAGACATCTCTCAAACCAGAATTAGGAGATTAAAGTCC
 ATGACATTAGACGTATTTACCAAGGTGGTTTCTCAAGCTGACTCCAGAGGCGAGTTTCCTGAGCAACGAACAACCTAGACGC
 1 M T L D V F T K V V S Q A D S R G E F L S N E Q L D A
 ATTGGCAAATGTTGTTAAAGAAGGCAACAAACGCTTGGATGTTGTTAAACCGCATCACCAGCAACGCTTCTGCGATCGTTA
 81 L A N V V K E G N K R L D V V N R I T S N A S A I V T
 CCAACGCTGCTCGTGCGTTGTTGAAGAACAACCCAGTTGATTGCTCCTGGTGGTAACGCTTACACCAACCGTCGCATG
 161 N A A R A L F E E Q P Q L I A P G G N A Y T N R R M
 GCTGCTTGTTTACGCGACATGGAAATCATCTTGCCTTACGTCACCTACGCTATCCTCGCAGGCGATGCTAGCGTTCTAGA
 241 A A C L R D M E I I L R Y V T Y A I L A G D A S V L D
 CGATCGCTGCTTGAACGGCTTGCAGCAACATACCAAGCTTTGGGAACCTCCTGGTTCTTCCGTAGCTGTTGGCGTTCAAA
 321 D R C L N G L R E T Y Q A L G T P G S S V A V G V Q K
 AAATGAAAGATGCTGCTGTTGGCATCGCTAACGACCCCAACGGAATCACCAAAGGTGATTGCAGCCAATTGATCTCTGAA
 401 M K D A A V G I A N D P N G I T K G D C S Q L I S E
 GTTGCTTGCTACTTTGATCGCGCTGCTGCTGTTGGTTAATTTTAACTAACAAGTCCAGCACCCACTGAAACCTAACTATC
 481 V A C Y F D R A A A V G *
 TTAGGTAACAGGTAAAAACTACGAAATTACTAAGGAGAATTTTACATCATGGTTAAAACCCCATACCGAAGCAATTGC
 561 M V K T P I T E A I A
 AGCTGCTGACACCCAAGGTCGTTTCTTAGGCAACACCGAACTCCAATCAGCTCGTGGTCGTTATGAGCGCGCTGCTGCTA
 641 A A D T Q G R F L G N T E L Q S A R G R Y E R A A A S
 GTTTAGAAGCTGCTCGTGGTTTACTTCCAACGCTCAACGCTTGATCGATGGTGCAACCCAAGCTGTTTACCAAAAATTC
 721 L E A A R G L T S N A Q R L I D G A T Q A V Y Q K F
 CCTTACACAACCCAAACCCCGGTCCTCAGTTGCTGCTGACAGCCGTGGTAAATCCAAGTGCCTGCTGACGTTGGTCA
 801 P Y T T Q T P G P Q F A A D S R G K S K C A R D V G H
 CTACCTACGCATCATCACCTATAGCTTAGTTGCTGGTGGTACTGGCCCTTGATGAATACCTAATTGCTGGTTTGGCTG
 881 Y L R I I T Y S L V A G G T G P L D E Y L I A G L A E
 AAATCAACAGCACCTTTGACCTATCTCCTAGCTGGTATGTTGAAGCTCTAAAGCACATCAAAGCTAATCATGGTTTAAGT
 961 I N S T F D L S P S W Y V E A L K H I K A N H G L S
 GGTCAAGCTGCTAACGAAGCTAACACCTACATCGACTACGCTATCAACGCTCTCAGCTAGATAGCTTCCTGCCCGAAAG
 1041 G Q A A N E A N T Y I D Y A I N A L S *
 GGGTGCAAGTGCTGGATGTAGTCATCATCTAGCAGTTGAATCTCGTTCCGGGCATAGTTTTATCTAGGAATAGCTGTCCG
 1121 TAGTTATAATCGGCAAAAAAATAGGGGAAACTTAAATGGCAATTACAACAGCAGCATCTAGGCTGGGGACAGAGCCTT
 1201 M A I T T A A S R L G T E P F
 TTAGCGACGCACCTAAAGTAGAATTACGCCCCAAAGCTAGCAGAGAAGAAGTAGAATCAGTGATTTCGCGCGTTTATCGG
 1281 S D A P K V E L R P K A S R E E V E S V I R A V Y R
 CACGTTTTGGGCAATGATTACATACTGGCATCAGAACGCCTTGTAAGTGCAGACTCTCTATTGAGAGATGGCAATCTGAC
 1361 H V L G N D Y I L A S E R L V S A E S L L R D G N L T
 AGTGCGTGAATTTGTGCGTAGCGTTGCTAAATCAGAACTCTACAAAAAGAAATTTTCTACAACAGCTTCCAAACTCGGC
 1441 V R E F V R S V A K S E L Y K K K F F Y N S F Q T R L
 TAATCGAACTTAACTATAAACACCTGTTAGGTGCGCGCCTTACGATGAGTCAGAAAGTTGTTTACCCTTGGACTTGATC
 1521 I E L N Y K H L L G R A P Y D E S E V V Y H L D L Y
 CAAAACAAAGGTTACGATGCCGAAATAGACTCCTATATAGATTCATGGGAGTATCAAAGCAATTTCCGGTGATAACGTCGT
 1601 Q N K G Y D A E I D S Y I D S W E Y Q S N F G D N V V

```

1681 TCCTTACTACCGTGGTTTTGAGACTCAAGTAGGACAAAAAACCGCAGGTTTCAACCGGATATTCCGTTTATACCGTGGTT
    P Y Y R G F E T Q V G Q K T A G F N R I F R L Y R G Y
1761 ATGCTAACAGCGATCGCGCTCAAGTAGAAGGTACAAAATCTCGTATTGGCGCGGTAATTTAGCTAGCAATAAAGCTTCT
    A N S D R A Q V E G T K S R I G A G N L A S N K A S
1841 ACAATTGTTGGGCCATCTGGAAGTAATGATAGTTGGGGTTTCCGGGCTTCAGCAGATGTCGCTCCCAAGAAAACTTGGG
    T I V G P S G T N D S W G F R A S A D V A P K K N L G

                                ↓
1921 TAATGCCGTTGGTGAAGGCGATCGCGTCTATCGACTAGAAAGTTACCGGAATCCGTAGCCCTGGCTACCCCAGTGACGGC
    N A V G E G D R V Y R L E V T G I R S P G Y P S V R R
2001 GGAGCAGCACGGTATTTATAGTACCTTACGAGCGCTCTCTGACAAGATCCAGCAAGTTCACAAGCAAGGTGGCAAAATC
    S S T V F I V P Y E R L S D K I Q Q V H K Q G G K I
2081 GTTAGCGTCACCTCTGCGTAAGATAAGCGGATTTGTGTCAGTGGTCAGTGGGTCATTAGTTAGTAGTCAGGATTCAAACAA
    V S V T S A *
2161 CTGACAACCTAGCAACTGACAACCTAACACGCGACAGCGCTATAAACTTACAACAATAGGAGATAGAGAAGTAATGTTTCGG
    M F G
2241 TCAAACCACACTTGGGGCTGGTAGCGTTTCTTCTCTGCTAGTCGAGTATTTGCTACGAAGTCGTAGGCTTGGCGCAAA
    Q T T L G A G S V S S S A S R V F R Y E V V G L R Q S
2321 GCTCAGAAACAGACAAAAACAAATACAACATCCGGAATAGCGGTAGTGTGTTTATCACAGTACCATACAGCCGGATGAAC
    S E T D K N K Y N I R N S G S V F I T V P Y S R M N
2401 GAAGAATACCAACGATTACCCGCTTGGGTGGCAAAATCGTCAAGATTGAGCAATTAGTTTCCGCAGAGGCATAATAGCG
    E E Y Q R I T R L G G K I V K I E Q L V S A E A * *
2481 CCCTGGTAATGATAGAACCCAGTGTGGAAGAATTTCCCGCAGAGAACGGGCCACAGCTAACACCAGAACTAGCCATAGCT
    M I E P S V E E F P A E N G P Q L T P E L A I A
2561 AATCTGCAATCATCAGACTTAAGTCTCCGCTATTATGCTGCTTGGTGGTTAGGTAAGTATCGGGTGAAAGAAAGTGCTGC
    N L Q S S D L S L R Y Y A A W W L G K Y R V K E S A A
2641 TGTGATGCTTTAATTGCGGCGTTAGAGGATGAAGCCGATAGAAGTGAAGTGGTGGTTATCCTTTGCGGCGTAACGCG
    V D A L I A A L E D E A D R T E L G G Y P L R R N A A
2721 CCAGAGCATTAGGGAAATGGGCAATCGTAAAGCCGTGCCAGGCTTAATTAAGTGGTGGAGTGTCTGACTTTTACGTG
    R A L G K L G N R K A V P G L I N C L E C P D F Y V
2801 CGTGAAGCAGCAGCCCAATCGCTGGAAATGCTCAAAGACAAAACAGCAGCACCAGCACTCATCAATTAAGTCTGATGGGG
    R E A A A Q S L E M L K D K T A A P A L I K L L D G G
2881 AGTCGCACAGGCCGTGCAAGTAACAGGGCGACCTCATTTAGTCCAACCTACGAAGCAGTATTAGAAGCGTTAGGAGCTA
    V A Q A V Q V T G R P H L V Q P Y E A V L E A L G A I
2961 TTGGTGCTACTGATGCCATCCCTCTGATTGAGCCATTTCTAGAGCATCCAGTCTCACGAGTGCAGTGCAGCCGCCGCTAGG
    G A T D A I P L I Q P F L E H P V S R V Q C A A A R
3041 GCAATGTACCAACTGACACAAGAACCAGTATATGGAGAGCTGTGGTAAAAGTGTAGCAGGTAACGACCTCAACCTACG
    A M Y Q L T Q E P V Y G E L L V K V L A G N D L N L R
3121 ACGCGTTGCTTTAGGTGACTTGGGTGCAATTGGGTACTTGGCAGCAGCAGAAGCGATCGCTAACGCCAAAGCCGAAAAA
    R V A L G D L G A I G Y L A A A E A I A N A K A E N S
3201 GCTTCAAACCTATTGCCCTCAAAGGATTGCTAGAACATCAGATGTCAGCAGAGTCAAACGCCCTCTCAATATCTGATCAA
    F K L I A L K G L L E H Q M S A E S N A L S I S D Q
3281 GCTATCCGGGTCATGAACCTTATGGATTGATTGTTGTAGTCAGTGGTCAGTAGTTAGTAGTCAGTAGCAACTGACAAC
    A I R V M N L M D S L L *
3361 ACCACTGGCGACTGACTATTGATTGATTGTGTACAGACGTTGTATACAACGCTCTACTAATGACTAATGAACCTAATTA

```

Fig. 3. Nucleotide sequence of the *Anabaena* PC operon. The PC β subunit extends from nucleotide 1 to nucleotide 506. The PC α subunit extends from nucleotide 609 to nucleotide 1097. The *cpcC* gene ORF extends from nucleotide 1238 to nucleotide 2098. The *cpcD* gene ORF extends from nucleotide 2233 to nucleotide 2472. The *cpcE* gene ORF extends from nucleotide 2489 to nucleotide 3316. Arrows indicate potential transcription initiation sites. Underlined sequences indicate potential -10 and -35 promoter recognition sequences. The standard single letter amino acid code is used.

fixed nitrogen) and uninduced (growth in the presence of ammonium chloride) cultures. Northern blots were prepared from these RNAs and probed with DNA containing PC and linker genes. The results are shown in Figure 2.

A probe prepared from the PC β (*cpcB*) coding sequence hybridizes to two RNA bands of 1.4 kb (transcript I) and 3.6 kb (transcript II), a pattern similar to that seen for the red light-inducible PC genes in *F. diplosiphon* (Conley *et al.*, 1985, 1986).

Fig. 4. Restriction and transcript map of the *Anabaena* 7120 PC operon. Arrows at the top of the figure indicate the approximate start and stop locations for transcripts I, II and III. The ORFs are indicated by boxes below the transcripts. The region of DNA sequenced is indicated by the shaded line beneath the ORF indicators. Restriction endonuclease sites within and flanking the operon are indicated *Hind*III (Δ), *Hinc*II ($\square$), *Clal* ($\bullet$), *Xba*I ($\blacktriangle$). Subclones used for analysis are indicated below the restriction sites. pAN410 contains the 4.3-kbp *Hinc*II fragment cloned into the *Hinc*II site of pUC18. pAN420 contains a 1.5-kbp *Hind*III–*Xba*I fragment cloned into the same sites in pUC18. The arrows at the bottom of the figure indicate the sequencing strategy.

[illegible][illegible]

Fig. 5. Comparison of PC β and α subunit amino acid sequences. The deduced amino acid sequences of the *Anabaena* 7120 (An) subunits are compared to the PC subunit sequences of *Synechococcus* 6301 (Sy) (Freidenreich *et al.*, 1978; Walsh *et al.*, 1980); *Mastigocladus laminosus* (Ma) (Frank *et al.*, 1978); and *Agmenellum quadruplicatum* (Ag) (Pilot and Fox, 1984; de Lorimier *et al.*, 1984). Blank positions indicate identity with the *Anabaena* sequence and dashes indicate gaps. Cysteine residues which are sites of chromophore attachment are indicated with (*). The notation is as in Figure 3.

		20		40
An	M A I T T A A S R L G T E P F S D A P	K V E L R P K A S R E E V E S V I R A V Y		
Ma	- A	N A I S D A N V		
MaPEC	- S - S V E A I K D E V - K I E N W S E D L Q I F K T A			
		60		80
An	R H V L G N D Y I L A S E R L V S A E	S L L R D G N L T V R E F V R S V A K S E		
Ma	E (end of sequence)			
MaPEC	E Q F (end of sequence)			
		100		120
An	L Y K K K F F Y N S F Q T R L I E L N Y K H L L G R A P Y D E S E V V Y H L D L			
		140		160
An	Y Q N K G Y D A E I D S Y I D S W E Y Q S N F G D N V V P Y Y R G F E T Q V G Q			
MaPEC			(Start sequence)	
		180		200
An	K T A G F N R I F R L Y R G Y A N S D R A Q V E G T K S - R I G A G N L A S N K A			
MaPEC	V E R F G - A S K I M - - -			
		220		240
An	S T I V G P S G T N D S W G F R A S A D V A P K K N L G N A V G E G D R V Y R L			
MaPEC	- P T S P I A A S T S S R T L V G M - - - - - - - - - I E A I			
		260		280
An	E V T G I R S P G Y P S V R R S S T V F I V P Y E R L S D K I Q Q V H K Q G G K			
MaPEC	A G L N T N - - - V A R Q Y T D A T Y E I R			
An	I V S V T S A			
MaPEC	K I P S			

Fig. 6. The amino acid sequence of the *cpcC* gene product the *Anabaena* 7120 (An) PC operon compared to the amino acid sequence of *M. laminosus* phycobilisome rod linkers. The deduced amino acid sequence of the third ORF of the operon (Figure 3) is compared to the partial sequences of the *M. laminosus* L_R^{34,PC} (Ma) and L_R^{34,PEC} (MaPEC) (Fuglistaller *et al.*, 1985). The notation is as in Figure 3.

A probe prepared from the coding region of the *cpcC* gene hybridized only to the 3.6-kb transcript. Finally, a probe prepared from the coding sequence of the *cpcE* gene hybridizes to both transcript II, and a second 1.4-kb transcript (transcript III). The sizes and approximate locations of the three transcripts are indicated in Figure 4. While these three transcripts are the dominant hybridizing bands, larger transcripts (5 and 9.5 kb) are also observed with all three probes. No significant difference in hybridization patterns is observed in the RNA from uninduced or induced cultures. Transcripts I and II are observed throughout the 36-h induction period (data not shown).

A probe prepared from the 600-bp *Hind*III–*Hinc*II fragment immediately downstream from the *cpcE* gene (Figure 4) fails to hybridize to transcripts II and III, but does hybridize to the longer transcripts (data not shown). These data suggest that partial read-through of the downstream terminator (Figure 9b) produces the longer transcripts.

Densitometer traces of the blots in Figure 2 reveal that transcript I is present in an ~20-fold higher concentration than transcripts II and III, consistent with the relative number of PC subunits compared to linker proteins in the phycobilisomes. In addition, if the assignment of the *cpcE* ORF to L_{RC}^{27,PC} is correct, the transcription of the operon as in Figure 4 ensures that

		20		40
An	M F G Q T T L G A G S V S S S A S R V F R Y E V V G L R O S S E T D K N K Y N I			
Ma	I D F M N E N			
		60		80
An	R N S G S V F I T V P Y S R M N E E Y Q R I T R L G G K I V K I E Q L V S A E A			
Ma	R Y N S M H F T R A G			

Fig. 7. The amino acid sequence of the *cpcD* gene product of the *Anabaena* 7120 (An) PC operon compared to the amino acid sequence of the *M. laminosus* phycobilisome rod-terminating linker. The deduced amino acid sequence of the fourth ORF in the operon (Figure 3) is compared to the amino acid sequence of the *M. laminosus* L_R^{8,9,PC} linker (Fuglistaller *et al.*, 1986a,b). Notation as in Figure 3.

there will be equal amounts of transcript coding for the two ends of the phycobilisome rods (*cpcD* and *cpcE*).

Mapping of the upstream transcription initiation site

The 5' ends of transcripts I and II were mapped by nuclease S1 digestion and primer extension. Nuclease S1 mapping is shown in Figure 10A. Single transcription start sites are observed in the RNA prepared either from induced or uninduced cultures, 243 bp upstream from the translation start for the β subunit (indicated with arrow). No significant ORFs were found within the transcript sequences 5' to the β translation start. Finer mapping of the transcription start, shown in Figure 10B, using RNA prepared from uninduced cells reveals a single family of bands.

Fig. 8. Amino acid sequence of the *cpcE* gene product of the *Anabaena* 7210 (An) PC operon compared to the amino acid sequences of the α and β subunits of *Anabaena* PC. The deduced amino acid sequence of the *cpcE* gene (An) is compared to the deduced amino acid sequences of the PC subunits (Figure 5). The single letter amino acid code is used as in Figure 3. Beneath the coding sequences of the fifth ORF, the partial sequences of the α (italic) and β (plain) PC subunits of *Anabaena* are presented, aligned to maximize homology. Conserved residues are indicated with dashes, conservative replacements with dots.

The first protected base is indicated in Figure 10B by an arrow. These data suggest that there is a single transcription initiation site for transcripts I and II. We cannot, however, rule out the possibility that the mapped 5' end of these transcripts is the result of an RNA processing event operating on a transcript initiated many kilobases upstream.

Structure of the 5' leader sequence of the upstream transcripts I and II

Primer extension, in addition to S1 nuclease digestion, was used to map the 5' end of transcripts I and II. A primer prepared from pAn420, extending from the *HincII* site to an *AluI* site 90 bp upstream (Figure 4), was annealed to RNA prepared from vegetative cells and extended, with the results shown in Figure 11.

Two major extended products are observed. The longer product reflects extension to near the 5' end of the transcripts mapped by S1 nuclease. The other reverse transcription product corresponds to a termination event within the S1-defined transcript. We considered the possibility that termination of reverse transcription occurred as a consequence of secondary structure within the RNA. Examination of the 5' leader sequence revealed the potential structure shown in Figure 9c, in which the major termination sites of the reverse transcriptase are indicated by arrows. The calculated ΔG of this structure is approximately -23 kcal/mol (Salser, 1977). These data suggest that the RNA coding for the PC subunits has stable stem and loop structures at both the 5' and 3' ends. The first nucleotide in the 5' secondary structure (-243 , Figure 9), represents the 5' end of transcripts I and II as mapped by S1 nuclease digestion (Figure 10B).

Mapping of the 5' end of transcript III

S1 nuclease mapping of the 1.4-kbp downstream transcript is shown in Figure 12. In contrast to the single start site for transcripts I and II, S1 nuclease digestion of DNA annealed to transcript III shows a family of bands beginning ~ 260 bp upstream from the translation initiation site for the *cpcD* gene (indicated by the arrow in Figure 12). These results indicate that the 5' end of the downstream transcript is 'ragged', with ends mapping to a region internal to the *cpcC* gene coding region (residues 240–254, Figure 6). In order to verify these 5' 'ragged' RNA ends, an oligonucleotide complementary to the transcript

Fig. 9. Schematic representation of possible secondary structures in the *Anabaena* PC operon. (a) The proposed terminator of transcript I (Figure 4), beginning 10 bp downstream from the *cpcA* gene translational termination codon. (b) The proposed terminator of transcript III (Figure 4). The first G in the structure represents the final nucleotide in the translational termination codon for the *cpcE* gene. (c) The proposed secondary structure in the 5' leader sequence of transcripts I and II (Figure 4). Numbers refer to the DNA sequence in Figure 3. Arrows in (c) indicate the major termination sites observed during reverse transcription (Figure 11) of the RNA.

30 bp downstream from the family of bands observed in the S1 nuclease digestion was synthesized. When annealed to RNA prepared from uninduced cells and extended with reverse transcriptase (Figure 13 lanes a,b,d), a family of termination sites was again observed. Some of the bands observed in the primer extension are clearly due to extension into transcript II. The others appear to indicate multiple 5' ends of the downstream transcript. As discussed below, the relative levels of transcripts II and III are altered by culture growth conditions. Growth in low light results in an increase of transcript II relative to transcript III (Figure 14). It can be seen that the intensities of several of the bands observed in Figure 13 are dependent on growth conditions (compare lanes a and b), and thus are due to the primer binding to transcript II and terminating at secondary structures. The intensities of the other bands are independent of light regime, which suggests that they arise from extension to 5' ends of transcript III. The arrow in Figure 13 indicates the position of a faint band in the extended lanes, which represents the longest extended product, the intensity of which is independent of light regime.

Regulation of PC operon transcripts by light intensity

The above data show that the *Anabaena* PC operon is encoded on three distinct transcripts. The most abundant codes for the PC subunits themselves, the translation products of which should be 6-fold more numerous than the products of the linker genes as dictated by the biliprotein/linker stoichiometry in the phycobilisome. Transcript II, which originates at the same promoter as the PC transcript, encodes three linker sequences in addition to the PC subunits. The biliprotein/linker stoichiometry could be at least partially regulated by altering the amount of these two transcripts. The presence of the third, downstream, transcript suggests that another level of regulation might exist.

Changing the ratio of the rod linker to rod-core linkers has been previously demonstrated to alter *Anabaena variabilis* phycobilisome rod assembly *in vitro* (Yu *et al.*, 1981). The length of PC rods formed *in vitro* varied inversely with the rod-core/rod linker ratio ($L_{RC}^{27,PC}/L_R^{32,PC}$). Thus, the three transcripts observed for the PC operon could allow phycobilisome

Fig. 10. S1 nuclease analysis of transcripts I and II of the PC operon. A 5' end-labeled 1.5-kbp *HincII*–*HindIII* fragment containing the amino terminus of PC β and 5' flanking sequences (Figure 4) was prepared as described in Materials and methods and used in the mapping experiments and for generation of sequencing tracks. (A) Input DNA annealed to RNA prepared from uninduced (lane a) or induced (lane b) *Anabaena* cultures, or to yeast tRNA (lane c) and digested with S1 nuclease. The arrow indicates the last base prior to the AUG translation start codon for the *cpcB* gene. (B) Input DNA annealed to RNA prepared from uninduced *Anabaena* cultures (lane a) or yeast tRNA (lane b) and digested with S1 nuclease. The arrow indicates the position of the first protected base.

rod length to be regulated at a transcriptional level. An increase in transcript II relative to transcript III would result in an increase in the rod/rod–core linker ratio, yielding longer phycobilisome rods. A mediator of such a regulatory process *in vivo* should be light intensity.

We therefore determined the effect of changing light intensities on the relative abundances of transcripts I, II and III. Northern blots were prepared from RNA isolated from *Anabaena* 7120 cultures growth at either high (H) or low (L) light intensity for 12 h (Figure 14). These blots were probed with nick-translated DNA coding for a variety of *Anabaena* genes. The data in Figure 14 indicate that the transcription of Rubisco is independent of light intensity. Similarly, the expression of the *psbB* gene, encoding the P680 protein of PSII, is independent of light intensity (data not shown). These data also show that the level of transcript II is lower in high light relative to both transcripts I and III. If generation of the 1.4-kb PC transcript is due to partial termination following the biliprotein coding sequences, this would imply that termination is increased at higher light intensity. This is consistent with the levels of transcript I, which in a lower exposure of the blot in Figure 14a, were observed to be higher at high light intensity than at low light intensity (data not shown).

The level of transcript III appears to be relatively insensitive

to light intensity (Figure 14). This insensitivity is also reflected in the mapping of the downstream transcript by primer extension (Figure 13). The intensities of bands associated with transcript III are relatively independent of the RNA used, while the bands due to extension along transcript II decrease with increasing light intensity.

Densitometer traces (data not shown) of the Northern blots in Figure 14 reveal that the ~10-fold increase in transcript II which occurs under low light is not associated with significant changes in the levels of transcript III. Total hybridizable message for the *cpcE* gene (transcripts II and III) (Figure 14b) increases ~5-fold in low light relative to high light conditions, while message hybridizable to *cpcB* (transcripts I and II) (Figure 14a) remains essentially constant. These results are consistent with transcript I arising from a partial termination event following the *cpcA* gene. In addition, the densitometer tracings are consistent with the model in which the synthesis of transcript III is directed by an independent, constitutive promoter within the *cpcC* gene. The model in which transcript III is a product of processing of transcript II would require that the increase in substrate for processing (transcript II) be coordinated with a concomitant decrease in processing activity under low light conditions.

The above data suggest that the primary effect of light quantity on the transcription of the PC operon is to regulate the relative

Fig. 11. Primer extension analysis of the 5' end of transcripts I and II of the PC operon. The 5' end-labeled primer which binds downstream from the translation start site of the PC β subunit was prepared as described in Materials and methods. Input DNA was annealed to 50 μ g of total *Anabaena* RNA prepared from an uninduced culture (lane a) or to yeast tRNA (lane b), and extended by addition of AMV reverse transcriptase. Sequencing tracks were generated from the 5' end-labeled fragment from Figure 10. The arrow indicates the translation start codon for the cpcB gene.

amounts of transcripts II and III. If a correct assignment of the cpcE ORF is assumed, then the $L_{RC}^{27.PC}/L_{RC}^{32.PC}$ ratio is transcriptionally regulated by light intensity, which would allow transcriptional regulation of phycobilisome rod length.

Discussion

Phycocyanin operon coding sequences

The major transcripts from the phycocyanin operon contain five

Fig. 12. S1 nuclease analysis of transcript III. A 5' end-labeled *Hind*III–*Xmn*I fragment internal to the cpcC coding region (Figure 4) was prepared as described in Materials and methods and used for mapping experiments and generation of sequencing tracks. Input DNA was annealed to 50 μ g of RNA prepared from an uninduced *Anabaena* culture (lanes a and b) or yeast tRNA (lane c) and subjected to S1 nuclease digestion (lanes a and c). The arrow indicates the translation start codon for the cpcD gene (Figure 4).

ORFs. The first two code for the β and α subunits of phycocyanin, followed by coding regions for two, and possibly three linker proteins. A similar structure is observed for the PC genes and linkers in *Synechococcus* 7002 (Bryant *et al.*, 1987) and the red light-induced PC genes in *F. diplosiphon* (Conley *et al.*, 1985, 1986). The relative orientation of the PC subunits, β followed by α , is conserved among the cyanobacteria. By contrast, the APC genes in both *Synechococcus* 7002 and *Synechococcus* 6301 have the opposite orientation (α followed by β) (Bryant *et al.*, 1985, Houmard *et al.*, 1987). Relatively large intercistronic sequences (~ 120 bp) are observed between coding regions in the *Anabaena* 7120 operon. This size has been observed in other polycistronic messages from cyanobacteria.

Fig. 13. Primer extension analysis of the 5' end of transcript III of the PC operon. An oligonucleotide (17-mer) complementary to the coding region of the *cpcD* gene (Figure 4) ~ 30 bp 3' to the transcript ends mapped in Figure 12 was synthesized, and annealed to RNA prepared from *Anabaena* cultures grown in high light (lane a), low light (lane b), or intermediate light intensity (lane d) used in Figure 12. The primer was also annealed to yeast tRNA (lane c). Annealed primers were extended by addition of AMV reverse transcriptase. Sequencing tracks were generated by annealing the primer to pAN410 and carrying out standard dideoxy sequencing reactions. The arrow indicates the position of a faint band which represents the longest reverse transcription product, the intensity of which is independent of light intensity.

The structure of PC from *Mastigocladus laminosus* has been determined by X-ray diffraction to 3 Å resolution (Schirmer *et al.*, 1985). These data have been used in conjunction with optical data to develop a model for energy transfer within assembled PC trimers (Mimuro *et al.*, 1986). As indicated in Figure 5, PC from *Anabaena* has a very high degree of homology to both the α (85%) and β (90%) subunits of *Mastigocladus laminosus*. Since most of the differences in primary structure between the two cyanobacteria are conservative amino acid replacements, the structure and energy transfer properties of the *Anabaena* PC should be homologous to those described for *M. laminosus*.

Sequence homologies between the biliproteins and linker polypeptides have been previously reported (Fuglistaller *et al.*, 1985). As similar homologies are observed among the biliprotein subunits (Troxler *et al.*, 1975), it has been suggested (Fuglistaller *et al.*, 1985) that all of the proteins of the phycobilisome are derived from a single ancestral progenitor. The sequences of the *Anabaena* PC operon coding regions are in agreement with these observations. The deduced amino acid

Fig. 14. Analysis of changes in transcript levels in response to light intensity. Northern blots of total *Anabaena* RNA prepared from uninduced cultures grown in high (H) and low (L) light were probed with (a) a 190-bp *HincII*–*XbaI* fragment internal to the *cpcB* gene and (b) a 850-bp *XbaI*–*HindIII* fragment containing the 3' end of the *cpcE* gene (Figure 4). Additionally, (c) a 1.4-bp *HindIII* fragment of pAN606 (Nierzwicki-Bauer *et al.*, 1984) coding for the small subunit of Rubisco of *Anabaena* 7120 was used as a probe. Transcript sizes were determined as in Figure 2.

sequences of the linker polypeptides of *Anabaena* have regions of homology with the PC subunits. The amino terminus of the rod linker (Figure 6) has more homology to the α subunit of PC than the β subunit (data not shown). This is in contrast to the sequence of the PC rod linker in *Mastigocladus* presented by Fuglistaller *et al.* (1985), where greater homology to the β subunit was indicated. The carboxy terminus of the *Anabaena* PC rod linker (Figure 6) shows good homology with the carboxy terminus of the 9-kD rod-capping polypeptide (Figure 7), again in agreement with previous observations (Fuglistaller *et al.*, 1985).

The deduced amino acid sequence of the *cpcE* gene product, the proposed rod-core linker (Figure 8), also has homology to the PC subunits. The amino and carboxy termini of this sequence have homology to the termini of the α subunit of PC. Internal regions of the sequence have homology to the β subunit. The *cpcE* gene product does not, however, appear to share the common carboxy terminal sequence found in the upstream polypeptides.

Structure and origins of transcriptional units

The PC operon of *Anabaena* is transcribed in three major RNAs. The three-transcript structure of the PC operon could arise from either defined synthesis of the transcripts or specific degradation events. The partial degradation of a single large transcript (transcript II) could result in the formation of transcripts I and III. The multiple transcripts could also arise from partial transcriptional termination following the *cpcA* gene generating transcripts I and II, combined with a second promoter within the *cpcC* gene generating transcript III. Finally, the multiple transcripts could arise from a combination of synthesis and specific degradation. We suggest that the second, two promoter-two terminator model, is the correct one.

The generation of the three transcripts by degradation of a single large transcript would require a specifically higher degree of degradation in the center of the transcript relative to both the 5' and 3' ends (Figure 4). The mechanism for this possibility is unclear. While the secondary structure following the *cpcA* gene (Figure 9a) suggests a possible RNA processing site, no similar secondary structure is found at the 5' end of transcript III. In the two promoter-two terminator model, a strong upstream promoter would direct the synthesis of both transcripts I and II, the shorter transcript (I) resulting from a partial termination event at the secondary structure following the biliprotein coding sequences (Figure 9a). The downstream transcript would then originate at its own promoter, internal to the *cpcC* gene. While the nature of the promoter sequence for *Anabaena* RNA polymerase has not been unequivocally determined, sequence homologies are observed 5' to the apparent transcription start sites for the upstream (I and II) (Figure 10B) and downstream (III) (Figure 13) transcripts (see underlined sequences in Figure 3). The source of the multiple 5' ends of transcript III is unclear, but similar 'ragged' 5' ends are observed in the transcript of T4 bacteriophage gene 23, a transcript which also originates at a promoter internal to the coding sequence of an upstream gene (Kassavetis and Geiduschek, 1982). These authors suggest that the multiple ends observed are due to processing or decay of the transcripts.

Finally, the three transcripts could arise from a combination of transcriptional termination and RNA processing. In this model, transcripts I and II are generated by partial read-through of the terminator following *cpcA*, and transcript III arises from degradation of transcript II. This model alleviates the requirement for a promoter internal to the *cpcC* coding region, and may explain

the origin of the 'ragged' ends of transcript III. However, this model would require that the RNA processing enzyme activity be light regulated, in that the level of transcript II is dependent on light intensity, while the level of transcript III is not.

The polycistronic *puf* operon in *Rhodobacter* has a structure which is similar to the PC operon described here. The *puf* operon contains four coding sequences, the first two encoding the structural genes for light-harvesting components, followed by the coding sequences for two much less abundant reaction center proteins (Belasco *et al.*, 1985). Two transcripts are synthesized from these coding sequences: a 2.6-kb transcript coding for all four polypeptides and a much more abundant 0.5-kb transcript which codes for only the light-harvesting components. In contrast to the *Anabaena* PC operon in which transcripts I and II have a common 5' end, the two *puf* transcripts have different 5' ends, beginning 75 and 104 bp upstream from the first structural gene for the 2.6- and 0.5-kb transcripts respectively (Zhu *et al.*, 1986).

The 5' leader sequence of transcripts I and II has a potentially stable secondary structure, which may serve to enhance the stability of the message in the cell. This may account for the stability of the PC transcripts during a 36-h induction period (Figure 2), relative to a wide variety of other vegetative cell transcripts which have been observed to decrease extensively after a similar induction (J.Lang, unpublished data). We cannot at this time rule out the possibility that the 5' secondary structure represents an RNA processing site.

Another very stable potential secondary structure is found following the α subunit of PC (Figure 9a). Termination of transcription at this structure would result in the formation of the 1.4-kb RNA observed to hybridize with probes prepared from the PC β subunit gene (Figure 2a). In the absence of termination at this structure, transcription could continue another 2.2 kb to the secondary structure following the *cpcE* gene (Figure 9b). Examination of the sequence 3' to the coding sequence of the PC α subunit of *Agmenellum quadruplicatum* (Pilot and Fox, 1984) indicates that a similar gene organization (PC α subunit followed by $L_R^{32,PC}$), and secondary structure, also exists in this cyanobacterium. Termination at this secondary structure in *Anabaena* could again be incomplete, resulting in the rare formation of longer transcripts. Again, the data presented here do not rule out the possibility that these secondary structures are processing sites. However, as discussed above, we propose that the downstream 1.4-kb RNA arises from a promoter within the coding sequence of the *cpcC* gene.

Transcriptional regulation of the PC operon

Regulation of photosynthetic genes by light has been observed in a variety of plants and cyanobacteria. In cyanobacteria, this regulation has been most extensively studied in chromatically adapting strains, such as *F. diplosiphon* (Conley *et al.*, 1985, 1986). White light-enhanced transcription from phycobiliprotein promoters has also been implied (Bryant *et al.*, 1987). We did not observe light-induced activation of the promoters of the PC operon (Figure 14a).

On the other hand, the data in Figure 14 suggest that light regulation of the PC operon in *Anabaena* occurs at the level of transcriptional termination. The two promoter-two terminator model suggests that under low light growth conditions, termination of transcription following the PC α subunit gene is decreased. As the level of transcript III is relatively insensitive to light intensity, we suggest that the downstream promoter is constitutive. If transcript III is a product of processing of transcript II, a light-dependent increase in processing would be required,

leading to enhanced levels of transcripts I and III under high light. However, in the data in Figure 14, the magnitude of the light-dependent change in transcript II is not reflected in changes in transcript III.

The final result of a light-dependent regulatory event would be expected to involve changes in efficiency of energy trapping or utilization. If our assignment of the *cpcE* ORF to $L_{RC}^{27,PC}$ is correct, and the levels of translated products of the operon are proportional to message levels, a higher $L_R^{32,PC}/L_{RC}^{27,PC}$ linker ratio would be expected in low light. This would result in the assembly of longer phycobilisome rods (Yu *et al.*, 1981), and more efficient light trapping.

Materials and methods

Materials

All chemicals used were of reagent grade. Restriction endonucleases, nick-translation kits, T4 polynucleotide kinase and S1 nuclease were obtained from Bethesda Research Laboratories. Alkaline phosphatase was obtained from Boehringer Mannheim. DNA polymerase I Klenow fragment, T4 DNA ligase, deoxynucleoside and dideoxynucleoside triphosphates were obtained from Pharmacia. Reverse transcriptase was obtained from BioRad. [γ - ^{32}P]dATP (3000 Ci/mmol), [α - ^{32}P]dCTP (3000 Ci/mmol) and [^{35}S]dATP α S (>600 Ci/mmol) were obtained from Amersham. Oligonucleotides were synthesized with an Applied Biosystems synthesizer.

Strains and culture conditions

Anabaena 7120 was grown in Kratz and Myers (1955) medium supplemented with 2.5 mM (NH₄)₂SO₄. Cultures were grown under fluorescent lights of three intensities of 20 (low light), 200 and 400 (high light) foot candles. *Escherichia coli* TB1 (Bethesda Research Laboratories) was used to maintain pUC18 and 19 (Norlander *et al.*, 1983) plasmid derivatives. *E. coli* JM105 (Yanisch-Perron *et al.*, 1985) was utilized for growth of M13 clones, and *E. coli* ED8767 was used to maintain the cosmid (pTBE) library of *Anabaena* 7120 DNA.

Isolation of DNA and RNA, Southern and Northern hybridizations

High mol. wt DNA was prepared from *Anabaena* as in Grosveld *et al.* (1982). RNA was isolated from *Anabaena* as described by Golden *et al.* (1987). Alkaline transfer of DNA to a nylon membrane (Reed and Mann, 1985) was utilized for preparation of Southern blots. For preparation of Northern blots, RNA was glyoxalated (McMaster and Carmichael, 1980) and subjected to agarose gel electrophoresis. Transfer of RNA to the nylon membrane (Gene Screen Plus), as well as hybridization conditions were as suggested by the manufacturers (Du Pont Company) except for low stringency hybridizations, for which hybridization and wash temperatures were decreased to 60°C.

Library construction and sequence analysis

A library of *Anabaena* 7120 DNA was prepared utilizing the cosmid vector pTBE (Grosveld *et al.*, 1982) containing *CpfI* inserts cloned into the *Bam*HI site of the vector. We are grateful to Dr J.W. Golden for the construction of this library. Screening of the cosmid bank using the *Cyanophora* probe was carried out as in Maniatis *et al.* (1982). The plasmids pUC 18 and 19 (Norlander *et al.*, 1983) were used for subcloning. The M13 vectors mp18 and 19 (Norlander *et al.*, 1983) were used for sequence analysis. The majority of the DNA sequence was obtained utilizing M13 sequencing templates prepared from the pAN410 and pAN420 *Anabaena* DNA. Sequencing templates contained either random inserts generated by ligation of sonicates of insert concatamers into the *Sma*I site of the vectors (Bankier and Barrell, 1983), or ordered deletions generated by *Bal*31 digestion (Maniatis *et al.*, 1982) of M13 replicative forms. The sequence of M13 templates was determined as in Sanger *et al.* (1977) using [^{35}S]dATP α S (Biggin *et al.*, 1983). The remainder of the DNA sequence was determined by plasmid sequencing (Hattori and Sakaki, 1986) using synthesized oligonucleotides as primers.

S1 nuclease and primer extension mapping

The 5' end of transcripts I and II (Figure 4) was mapped by S1 nuclease digestion using the 1.5-kbp *Hind*III/*Hinc*II insert from pAN420. The plasmid was digested with *Hinc*II, treated with alkaline phosphatase and end labeled with polynucleotide kinase ([γ - ^{32}P]ATP). The labeled plasmid was then digested with *Hind*III and the 1.5-kbp fragment isolated. The fragment was annealed to 50 μ g *Anabaena* RNA and digested with S1 nuclease (Sharp *et al.*, 1980). Sequencing tracks were also generated from the isolated fragment (Maxam and Gilbert, 1980). The same procedure was used for mapping the 5' end of transcript III (Figure 4), using a 680-bp *Xmn*I/*Hind*III fragment from pAN410 (Figure 4). The 5' end of transcripts I and II was mapped by primer extension utilizing a primer generated from

pAN420. The plasmid was digested with *Hinc*II and end labeled as above, followed by digestion with *Alu*I. A 90-bp *Alu*I/*Hinc*II fragment was isolated, followed by annealing to 50 μ g *Anabaena* RNA and extension by AMV reverse transcriptase under conditions suggested by the enzyme supplier (Bio-Rad). Sequencing tracks were the same as used in the S1 mapping of these transcripts for transcripts I and II. Primer extension mapping of transcript III (Figure 4) was carried out using a synthesized oligonucleotide (17 mer) (5'-CCGCTCGTAAGGTACTA-3') as a primer. The primer was end labeled and annealed to 50 μ g of RNA by heating to 90°C for 1 min and cooling in an ice bath. The primer was extended as before. Sequence tracks were generated by annealing the unlabeled primer to pAN410, and extending in the presence of dideoxy nucleotides (Hattori and Sakaki, 1986).

Acknowledgements

We are grateful to P. Lemaux for the *Cyanophora* probe. W.R.B. is a postdoctoral trainee in Developmental Biology (NIH grant HD07136). This work was supported by grants from the NIH (GM 21823) and the Monsanto Co.

References

- Bankier, A.T. and Barrell, B.G. (1983) *Nucleic Acid Biochemistry*. Elsevier Scientific Publishers, Ireland, **B508**, 1–34.
- Belasco, J.G., Beatty, J.T., Adams, C.W., von Gabain, A. and Cohen, S.N. (1985) *Cell*, **40**, 171–181.
- Bogorad, L. (1975) *Annu. Rev. Plant Physiol.*, **26**, 369–401.
- Bryant, D.A., Guglielmi, G., Tandeau de Marsac, N., Castets, A.-M. and Cohen-Bazire, G. (1979) *Arch. Microbiol.*, **123**, 113–127.
- Bryant, D.A., de Lorimier, R., Lambert, D.H., Dubbs, J.M., Stirewalt, V.L., Stevens, S.E., Jr (1987) In Biggins, J. (ed.) *Prog. in Photosynthesis Research*, Vol. 4, Martinus Nijhoff, Dordrecht, pp. 749–755.
- Bryant, D.A., de Lorimier, R., Guglielmi, G., Stirewalt, V.L., Cantrell, A. and Stevens, S.E., Jr (1987) Proceedings of the VI International Congress of Photosynthesis, in press.
- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
- Conley, P.B., Lemaux, P.G. and Grossman, A.R. (1985) *Science*, **230**, 550–553.
- Conley, P.B., Lemaux, P.G. and Grossman, A.R. (1986) *Proc. Natl. Acad. Sci. USA*, **81**, 3924–3928.
- de Lorimier, R., Bryant, D.A., Porter, R.D., Liu, W.-Y., Jay, E. and Stevens, S.E., Jr (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7946–7950.
- Fisher, R., Tuli, R. and Haselkorn, R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3393–3397.
- Frank, G., Sidler, W., Widmer, H. and Zuber, H. (1978) *Hoppe-Seyler's Z. Physiol. Chem.*, **359**, 1491–1507.
- Freidenreich, P., Apell, G.S. and Glazer, A.N. (1978) *J. Biol. Chem.*, **253**, 212–219.
- Fuglistaller, P., Suter, F. and Zuber, H. (1985) *Biol. Chem. Hoppe-Seyler*, **366**, 993–1001.
- Fuglistaller, P., Suter, F. and Zuber, H. (1986a) *Biol. Chem. Hoppe-Seyler*, **367**, 601–614.
- Fuglistaller, P., Suter, F. and Zuber, H. (1986b) *Biol. Chem. Hoppe-Seyler*, **367**, 615–626.
- Gantt, E. (1981) *Annu. Rev. Plant Physiol.*, **32**, 327–347.
- Gasparich, G.E., Buzby, J., Bryant, D.A., Porter, R.D. and Stevens, S.E. (1987) In Biggins, J. (ed.) *Prog. in Photosynthesis Research*, Vol. 4, Martinus Nijhoff, Dordrecht, pp. 761–764.
- Glazer, A.N. (1981) In Stumpf, P.F. and Conn, E.E. (eds), *The Biochemistry of Plants*. Academic Press, New York, Vol. 8, pp. 51–96.
- Glazer, A.N. (1985) *Annu. Rev. Biophys. Chem.*, **14**, 47–77.
- Golden, S.S., Brusslan, J. and Haselkorn, R. (1987) *Methods Enzymol.*, in press.
- Grossman, A.R., Lemaux, P.G. and Conley, P.B. (1986) *Photochem. Photobiol.*, **44**, 827–837.
- Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M. and Flavell, R.A. (1982) *Nucleic Acids Res.*, **10**, 6715–6732.
- Haselkorn, R. (1978) *Annu. Rev. Plant Physiol.*, **29**, 319–344.
- Hattori, M. and Sakaki, Y. (1986) *Anal. Biochem.*, **152**, 232–238.
- Houmard, J., Maxel, D., Moguet, C., Bryant, D.A. and Tandeau de Marsac, N. (1986) *Mol. Gen. Genet.*, **205**, 404–410.
- Jahn, W., Steinbiss, J. and Zetsche, K. (1984) *Planta*, **161**, 536–539.
- Kassavetis, G.A. and Geiduschek, E.P. (1982) *EMBO J.*, **1**, 107–114.
- Kratz, W.A. and Myers, J. (1955) *Am. J. Bot.*, **42**, 282–287.
- Lemaux, P.G. and Grossman, A.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4100–4104.
- Lemaux, P.G. and Grossman, A.R. (1985) *EMBO J.*, **4**, 1911–1919.
- Lundell, D.J., Yamanaka, G. and Glazer, A.N. (1981) *J. Cell Biol.*, **91**, 315–319.
- Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.

- Mimuro,M., Fuglistaller,P., Rumbell,R. and Zuber,H. (1986) *Biochim. Biophys. Acta*, **848**, 155–166.
- Nierzwicki-Bauer,S.A., Curtis,S.E. and Haselkorn,R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5961–5965.
- Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
- Oquist,G. (1974a) *Physiol. Plant.*, **30**, 38–44.
- Oquist,G. (1974b) *Physiol. Plant.*, **30**, 45–48.
- Platt,T. (1986) *Annu. Rev. Biochem.*, **55**, 339–872.
- Pilot,T.J. and Fox,J.L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6983–6987.
- Peterson,R.B., Dolan,E., Calvert,H.E. and Ke,B. (1981) *Biochim. Biophys. Acta*, **634**, 237–248.
- Porter,G., Tredwell,C.J., Searle,G.F.W. and Barber,J. (1978) *Biochim. Biophys. Acta*, **501**, 532–545.
- Reed,K.C. and Mann,D.A. (1985) *Nucleic Acids Res.*, **13**, 7207–7221.
- Salsler,W. (1977) *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. 77, pp. 985–1010.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Schirmer,T., Bode,W., Huber,R., Sidler,W. and Zuber,H. (1985) *J. Mol. Biol.*, **185**, 257–277.
- Searle,G.F.W., Barber,J., Porter,G. and Tredwell,C.J. (1978) *Biochim. Biophys. Acta*, **501**, 246–256.
- Sharp,P.A., Berk,A.J. and Berget,S.M. (1980) *Methods Enzymol.*, **65**, 750–768.
- Tandeau de Marsac,N. and Cohen-Bazire,G. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1635–1639.
- Troxler,R.F., Foster,J.A., Brown,A.S. and Franzblau,C. (1975) *Biochemistry*, **14**, 268–274.
- Walsh,R.G., Wingfield,P., Glazer,A.N. and DeLange,R.J. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, **39**, 2060.
- Wood,N.B. and Haselkorn,R. (1980) *J. Bacteriol.*, **141**, 1375–1385.
- Yamanaka,G. and Glazer,A.N. (1983) In Papageorgiou,G.C. and Packer,L. (eds), *Photosynthetic Prokaryotes*. Elsevier, Amsterdam, pp. 69–90.
- Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
- Yu,M.-H. and Glazer,A.N. (1982) *J. Biol. Chem.*, **257**, 3429–3433.
- Yu,M.-H., Glazer,A.N. and Williams,R.C. (1981) *J. Biol. Chem.*, **256**, 13130–13136.
- Zhu,Y.S., Kiley,P.J., Donohue,T.J. and Kaplan,S. (1986) *J. Biol. Chem.*, **261**, 10366–10374.

Received on December 19, 1986