Light-dependent attenuation of phycoerythrin gene expression reveals convergent evolution of green light sensing in cyanobacteria

Ryan P. Bezy, Lisa Wiltbank, and David M. Kehoe

*Department of Biology, Indiana University, Bloomington, IN 47405; †Department of Natural and Applied Sciences, Mount Mercy University, Cedar Rapids, IA 52402; and ‡Indiana Molecular Biology Institute, Indiana University, Bloomington, IN 47405

Edited by Susan S. Golden, University of California at San Diego, La Jolla, CA, and approved October 4, 2011 (received for review May 10, 2011)

The colorful process of chromatic acclimation allows many cyanobacteria to change their pigmentation in response to ambient light color changes. In red light, cells produce red-absorbing phycocyanin (PC), whereas in green light, green-absorbing phycoerythrin (PE) is made. Controlling these pigments increases fitness by optimizing photosynthetic activity in different light color environments. The light color sensory system controlling PC expression is well understood, but PE regulation has not been resolved. In the filamentous cyanobacterium *Fremyella diplosiphon* UTEX 481, two systems control PE synthesis in response to light color. The first is the Rca pathway, a two-component system controlled by a phytochrome-class photoreceptor, which transcriptionally represses *cpeCDESTR* (*cpeC*) expression during growth in red light. The second is the Cgi pathway, which has not been characterized. We determined that the Cgi system also regulates PE synthesis by repressing *cpeC* expression in red light, but acts posttranscriptionally, requiring the region upstream of the *CpeC* translation start codon. *CpeC* RNA stability was comparable in *F. diplosiphon* cells grown in red and green light, and a short transcript that included the 5' region of *cpeC* was detected, suggesting that the Cgi system operates by transcript attenuation. The roles of four predicted stem-loop structures within the 5' region of *cpeC* RNA were analyzed. The putative stem-loop 31 nucleotides upstream of the translation start site was required for Cgi system function. Thus, the Cgi system appears to be a unique type of signal transduction pathway in which the attenuation of *cpeC* transcription is regulated by light color.

Photosynthetic gene expression is precisely regulated in response to environmental conditions and controlled at the transcriptional and posttranscriptional levels in plants, algae, and cyanobacteria (1–3). In some cyanobacteria, genes encoding photosynthetic light-harvesting proteins are regulated by light color by chromatic acclimation (CA) (4–6). CA-mediated accumulation of these proteins, which contain covalently attached bilin chromophores, allows the tailoring of the absorption profile of these structures to match the spectral distribution of ambient light. These changes maximize photon capture for photosynthesis, providing a selective advantage in changing light color environments (7). Two forms of CA, type 2 and type 3, exist in species that contain the two light-harvesting proteins phycocyanin (PC), which maximally absorbs red light, and phycoerythrin (PE), which maximally absorbs green light (8). CA is widespread, as nearly 75% of the species containing PC and PE are capable of one of these two types of CA (8). CA2-capable cyanobacteria produce more PE in green light than in red light but do not alter PC levels in response to light color. CA3-capable species regulate PE production in response to light color similarly to CA2 species, but also make more PE in red light than in green light.

Two signaling pathways control CA3 in *Fremyella diplosiphon* (6). The Rca two-component system transcriptionally regulates PC and PE production by repressing PE-encoding genes and activating PC-encoding genes in red light. The Rca sensor is the photoreceptor RcaE, the founding member of the cyanobacteriochromes (9–11). It contains a histidine kinase domain that modulates the activity of the OmpR-class transcription factor RcaC, which binds to direct DNA repeats upstream of CA3-regulated genes called the L box, activating red light expressed genes and repressing green light expressed genes (12–14). The Cgi system controls only the activity of genes that are highly expressed in green light, and may operate by controlling the *cpeCDESTR* (*cpeC*) operon (6), which encodes light-harvesting antenna proteins and a CA3 activator that regulates additional green light expressed genes (15, 16).

The pattern of gene regulation by the Rca and Cgi systems led to the proposal that CA2 and CA3 capabilities have evolved by loss of the Rca system from a CA3 species to create a CA2 species or by acquisition of the Rca system by a CA2 species to create a CA3 species (6, 17). This proposal was explored using the CA2 species *Nostoc punctiforme*, where sensor kinase/response regulator encoding genes were found adjacent to *cpeCGRI* (18). These proteins, CcaS/R, are highly similar to RcaE and RcaC but appear to function in an opposite manner in red and green light. It was proposed that the *F. diplosiphon* Cgi system, which is uncharacterized at the molecular level, is equivalent to the CcaS/CcaR transcriptional regulatory system. It was also proposed that CA3 capability evolved through CcaSR duplication to create the Rca system, which then diverged to transcriptionally regulate both PC- and PE-encoding genes.

In the descendants of endosymbiotic cyanobacteria, chloroplasts, posttranscriptional processes play an important role in regulating gene expression (19–23), often through the mRNA 5' leader. Stem–loop structures located near the translation initiation site are frequently used as binding sites for multiple proteins that influence the translation rate (22, 24). In cyanobacteria, posttranscriptional regulation of photosynthetic gene expression is also an important control mechanism, but is only known to occur via changes in mRNA stability (25–28).

Here, we characterize the molecular basis of the Cgi system regulation of *cpeC* in *F. diplosiphon*, finding that this system operates at the posttranscriptional level and, unlike known posttranscriptional regulation in cyanobacteria, does not involve differential mRNA stability. The region of *cpeC* upstream of the CpeC translation start site and downstream of the *cpeC* transcription start site (the “5' leader region”) contains a sequence

Author contributions: R.P.B., L.W., and D.M.K. designed research; R.P.B. and L.W. performed research; R.P.B., L.W., and D.M.K. analyzed data; and R.P.B., L.W., and D.M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: dkehoe@indiana.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1107427108/-/DCSupplemental.
required for Cgi control with predicted secondary structure and location that is similar to cis elements that regulate gene expression in plant and green algal chloroplasts. A short transcript at the 5′ end of cpeC was also detected. These findings indicate that the Cgi pathway is a transcription attenuation system that is regulated by light color and provide insights into the evolution of light regulated signal transduction pathways in CA2- and CA3-capable cyanobacteria.

**Results**

To identify cpeC cis element(s) needed for Cgi system operation and determine whether this pathway functions independently of the Rca system, several translational fusions were created using the cpeC upstream region, the non-light–regulated cpc1 5′ leader (29, 30), and the gusA reporter gene (31) (Fig. 1A). cpeC upstream and 5′ leader regions (pRB7) conferred a 10-fold increase in β-glucuronidase (GUS) activity in green versus red light (Fig. 1B), mirroring the previously measured 10-fold change in cpeC mRNA levels (15, 32–34). Thus, all elements needed for CA3 regulation of cpeC expression are contained in the region from −412 to +196 in pRB7. Replacing the L box direct repeat (pRB8) increased GUS activity levels in red light but not green light (Fig. 1B), confirming the Rca system’s repressing effect on cpeC expression in red light (14). The ∼2.5-fold green light induction remaining after L box replacement, identical to that measured for cpeC mRNA in rcaC and rcaE mutants (15, 33), is due to the Cgi system. This demonstrates that the Cgi pathway requires the upstream and/or 5′ leader regions of cpeC. Replacing the cpeC 5′ leader with the cpc1 5′ leader in pRB7 to create pRB1 also led to increased GUS activity in red light but only a slight increase in green light (Fig. 1B). This ∼2.5-fold green light induction was due to Rca regulation, since cpeC light color regulation was completely lost after the L box was replaced in pRB1 (pRB6) (Fig. 1B). These results demonstrate that although the Cgi system operates independently of the Rca system, it provides additional repression of cpeC expression in red light. They also show that the Cgi system requires the cpeC 5′ leader but not the region upstream of the transcription start site, suggesting that unlike the Rca system, the Cgi pathway acts post-translationally.

We measured rates of cpeC mRNA loss during growth in red and green light to determine if differential abundance of this transcript could be explained by light color-dependent mRNA degradation activity. An rcaE null mutant (11) was used for this experiment to eliminate the influence of the Rca system. There was no significant difference in cpeC mRNA stability during growth in red versus green light (red light half-life = 14.1 min, green light half-life = 16.3 min, P = 0.523) (Fig. 2). Thus, changes in cpeC mRNA levels mediated by the Cgi pathway are not due to different RNA degradation rates in red and green light.

The Cgi system does not appear to operate at the transcriptional level or the level of mRNA stability. We examined the possibility that it works by transcriptional attenuation, which should produce a short species of cpeC mRNA encompassing the 5′ leader region. Analysis of cpeC expression using a probe made to the coding region of the first gene in this operon revealed two transcripts, of ∼2.2 and 3.2 kilonucleotides (knt), that correspond to two- and three-gene transcription units (“long transcripts”) (Fig. 3A and B). These RNAs are 10 times more abundant in green light than red light in wild-type (WT) cells and 2.5 times more abundant under the same light conditions in an rcaC null mutant, due primarily to greater expression in red light (Fig. 3A and C), as previously noted (34). However, when a probe made to the cpeC 5′ leader region was used (Fig. 3B), an additional RNA (“short transcript”) was detected that was much smaller than the long transcripts (Fig. 3A Center). Polyacrylamide gel electrophoresis was used to size the short transcript at ∼125 nt (Fig. 3A Right). Quantification of the abundance of the long transcripts detected using the cpeC 5′ leader probe in WT and the rcaC mutant revealed patterns of expression in red and green light comparable to those obtained when using a probe from within the coding sequence, with a 10-fold difference for WT and a ∼2.5-fold difference for the rcaC mutant (Fig. 3C) (34). In WT, the short transcripts were approximately three times more abundant in green light than red light.

---

Fig. 1. Role of the cpeC 5′ leader in Cgi regulation. (A) Diagrams of the constructs used. Upstream and 5′ leader regions of cpeC with (pRB7) and without (pRB8) the L box (black triangle) were joined translationally to gusA. The CpeC start codon begins at +188, so these translational fusions include the first three amino acids of CpeC. The region upstream of cpeC with and without the L box, joined to the cpc1 5′ leader, was also joined translationally to gusA (pRB1 and pRB6). Bent arrows, transcription start site. (B) Relative mean rates of GUS activity from cell lysates of F. diplosiphon transformed with the indicated plasmids and grown in red light (RL) or green light (GL). The mean value (302 nmol of product per mg of protein per min) derived from cells transformed with pRB1 and grown in green light was set at 100%. At least five independently transformed lines were tested for each plasmid and light condition. Error bars show SE.

Fig. 2. Rates of cpeC RNA loss in F. diplosiphon during growth in red and green light. RNA blot analyses measured cpeC RNA levels in the rcaE mutant of F. diplosiphon during growth in red light (open circles) or green light (filled circles). Rifampicin was added at time 0. For each light condition, RNA levels are expressed as a percentage of cpeC transcripts at the zero time point, which was set to 100%. Means of all values were calculated after loading normalization using ribosomal values. A one-phase decay model was used for the nonlinear fit in the regression analysis (Prism 5, Graphpad Software). Six independent assays were conducted for each light condition. Error bars show SE.
light. However, the abundance ratios of long to short transcripts were different in red versus green light. In red light, it was 0.86, whereas in green light it was 3.6. Thus, there was a higher percentage of short transcripts in red light than in green light. The lack of a functional Rca system in the rcaC mutant dramatically affected the overall amounts of small transcripts in red and green light. The small RNA was six times more abundant in this mutant than in WT in red light, but only slightly lower than in WT in green light. The ratios of long to short transcripts in red and green light in this mutant were quite similar to those measured for WT, with values of 0.64 in red light and 3.0 in green light. These data suggest that the Cgi system has a role in shifting the ratio of short to long cpeC transcripts in red light versus green light, with a greater percentage of short transcripts in red light and a greater percentage of long transcripts in green light. These shifts are consistent with a Cgi regulatory mechanism of cpeC transcription attenuation in which red light causes more frequent premature termination and less accumulation of the long forms of this RNA.

Mfold RNA secondary structure analyses (35) of the cpeC 5′ leader predicted the formation of four stem–loop structures (Fig. 4A). To test their role in the Cgi system, deletion constructs

Fig. 3. Detection of a cpeC 5′ leader transcript. (A) Representative autoradiographs of RNA blot analyses for WT and the rcaC mutant in red light (RL) or green light (GL) after hybridization with a cpeC coding sequence probe (Left) or a probe encompassing the first 150 bp of the cpeC 5′ leader (Center and Right). The short transcript was resized using a polyacrylamide gel (right). Ribosomal loading controls (ribo) are shown. Boxed areas demark regions used to quantify the results. (B) Diagrams of the cpe operon, probe locations, and transcripts detected. Sizes and classifications as long or short transcripts are shown, as well as the genes included in the long transcripts. The ~5.2-knt transcript is detectable after a longer exposure (Fig. S1). Dashed lines, probe locations; bent arrow, transcription start. (C) Relative mean accumulation levels of short and long cpeC transcripts [boxed areas in (A)] in WT and rcaC mutant cells grown in red or green light. Values are expressed as a percentage of the long transcripts value for WT cells grown in green light after ribosomal normalization (ribo). Numbers in parentheses are the ratios (to the nearest tenth) of long to short transcripts for each light condition and strain. Data are from three independent RNA blot analyses. Error bars show SE.
were created that eliminated each putative stem–loop (stem1 to stem4), using pRB8 as the base construct (Fig. 4B), which lacks the L box and removes Rca system influence. The putative stem–loop sequences were eliminated rather than replaced to avoid creating alternative secondary structures. Stem1 included regions 1, 2, and 3 of the 5′ leader, whereas stem2, stem3, and stem4 corresponded to regions 2, 3, and 4 in Fig. 4A. Cells carrying these plasmids were analyzed for increased GUS activity during growth in red and green light. Only the removal of stem4 led to elevated GUS activity in red light (Fig. 4C), the same phenotype as obtained by replacing the entire cpeC 5′ leader region (Fig. 1). We tested whether this effect was specific for stem4 or due to its closeness to the putative Shine–Dalgarno (SD) region (arrow, Fig. 4A) by substituting random sequence for the region between stem4 and the SD region (pRB22) and measuring GUS activity in red and green light. This construct repressed GUS activity in red light as well as pRB8 (Fig. 4C), demonstrating that the effect of stem4 removal was not due to its proximity to the SD region. Thus, the stem4 region of the cpeC 5′ leader is important for Cgi system function.

Four modified forms of the 30-bp stem4 region were tested to determine the effect of the sequence changes in the stem while maintaining (pRB21) or reducing (pRB19) the free energy of folding, changing the loop sequences (pRB18), and modifying a single-nucleotide stem mismatch (pRB20) (Fig. 5A). Each was tested for Cgi system function in red and green light, with pRB8 and the stem4 deletion construct used as controls (Fig. 5B). These all resulted in the loss of red light repression of GUS activity. Thus, minor changes in the sequence and/or structure of stem4 disables the Cgi system, independent of the free energy of folding or whether they are located in the putative stem or loop region. Also, the position of stem4 is critical, since moving stem4 closer to or further from translation start eliminated Cgi system function (Fig. S2).

Discussion

Light color regulation of cpeC expression by the Cgi system in *F. diplosiphon* operates by posttranscriptional repression in red light. This regulation is not due to differential cpeC RNA stability in red and green light, but appears to operate via transcriptional attenuation. In addition, this system requires the cpeC 5′ leader region adjacent to the putative SD sequence, which is predicted to form a stem–loop structure. Thus, the Cgi system is a previously unidentified type of light regulated signal transduction pathway that appears to use a stem–loop structure adjacent to the translation start site, similar to those regulating the expression of many chloroplast genes (21–24).

Control of cpeC mRNA levels through a mechanism not involving differential RNA stability (Fig. 2) is unique among cyanobacterial genes that are posttranscriptionally regulated. Two mechanisms of posttranscriptional control of gene expression are known in cyanobacteria: light-regulated changes in the stability of the transcripts from *psbA* genes (25, 28, 36–39), which may operate via ribosome pausing rather than at the level of translation initiation (26, 40, 41), and antisense RNA regulation of *isi*A expression, which has been proposed to alter the degradation rates of transcripts from this gene (42). Postranscriptional control of *hlhA* and *hspA* gene expression has also been reported, but the mechanism(s) through which these processes operate are not known (27, 43).

The involvement and location of the stem4 region suggests that similarities may exist between the Cgi system and those regulating chloroplast translation in plants and algae, in which 5′ leaders of chloroplast mRNAs contain regulatory sequences near translation initiation sites (44–50). Although these sequences operate in different ways (22, 23), they often form stem–loop structures similar in size and distance from the translation start site as stem4 of the *cpeC* 5′ leader. Proteins, some of which are light and redox regulated, interact with these stem–loop structures and activate or repress translation rates of the transcripts. For example, the *rps7* and *psbA 5′ leaders* have protein binding sites within stem–loop structures upstream of their translation start codons, and these are 70–84% A–U rich, equivalent to the 80% A–U composition of stem4 in the *cpeC 5′ leader*. Although these structural similarities suggest the possibility of related mechanisms, the final responses differ. Transcription and translation are coupled in bacteria, so *cpeC* transcriptional attenuation may be regulated by this stem–loop. In chloroplasts, these processes are predominantly uncoupled (21), and these stem–loop structures are important in translational regulation. Overall, our data suggest that the Cgi system mechanism may be structurally related to a subset of mechanisms that provide an important form of regulation of chloroplast gene expression in plants and green algae.

The process through which the Cgi system might cause more frequent *cpeC* transcription attenuation in red light than green light is not yet clear, although it seems relatively inefficient because significant amounts of short transcript are still present in green light (Fig. 3C). It is unlikely to involve a riboswitch, because these require physical or thermal inputs (51, 52) that could not be provided by an elicitor such as light. It is also probably not via antisense RNA, because a single nucleotide change within stem4 abolished the Cgi repression of *cpeC* (Fig. 5). It is
also unlikely that this system represses cpeC expression by translating an ORF within the cpeC 5' leader, because introducing a stop codon within each of the three small ORFs within this leader failed to cause a major increase in reporter gene expression in red light (Fig. S3). The strong effect of any stem4 modification suggests that it may interact with one or more proteins, and its proximity to the SD and translation start codon suggest that either stem4 or its associated components may interact with the ribosome. Because transcription attenuation systems are controlled by varying the extent of coupling between transcription and translation, for the Cgi regulation of cpeC expression, attenuation should occur within the first coding region in the cpeC operon. Because the cpeC 5' leader is 196 nt long, the attenuated transcript should be longer than this length. However, the species detected from this region on the RNA blot was only ~125 nt long (Fig. 5), apparently ending at the 3' end of stem1 (Fig. 4). This size discrepancy might result from exonucleolysis of the 3' end of the attenuated transcript up to where the stem1 secondary structure would stop any further nucleolysis. This possibility is supported by 3' RACE results using RNA from a rcaE null mutant (34), which produced one predominant band that, when sequenced, proved to be cpeC sequence extending 413 nt from transcription start, or 226 nt from translation start (Fig. S4). The 3' end sequences were heterogeneous, suggesting that 3' CCA RNAs with CA3 different 3' ends were present in the sample. It is also possible that the proposed RNA stem-loop instead forms in the DNA; we do not currently have evidence to support either possibility. However, it is unlikely that a DNA structure is controlling transcription attenuation because if it did form in the DNA, the short cpeC transcript would likely extend before the structure, perhaps close to +30 (Fig. 4b).

Recently, it was proposed that orthologs of the N. punctiforme CcaSR proteins, which control CA2 regulation of PE expression, might make up the Cgi system in CA3-capable species (18). This is unlikely because the CcaSR system apparently acts transcriptionally, whereas the results presented here demonstrate that the Cgi system operates posttranscriptionally. It is conceivable that components such as CcaS and CcaR operate in the initial steps of the Cgi pathway, although this is doubtful because genes encoding these components are not present in the genome of another CA3-capable species, Synechococcus sp. PCC 7335 (18).

In addition, the differences we have found between the CcaSR and Cgi systems do not support the hypothesis that CA2 and CA3 regulatory systems are related by either the straightforward addition or subtraction of the Rca system (6, 17). They also demonstrate that CA regulation of PE expression must have evolved more than one time and in more than one way. This finding contrasts with our current understanding of the evolution of CA regulation of PC synthesis, which uses highly conserved Rca system components and operates through the L box regulatory element in all species examined thus far (30, 53). Whether the differences between the evolution of PE and PC regulation are the result of more recent evolution of CA control of PC expression, or of greater selection pressure acting on the Rca system, remains to be determined.

The original description of the CA2 versus CA3 phenotype noted that no CA2 species was able to completely halt PE production in green light, unlike a number of CA3 species (8). A molecular explanation for this difference is now possible. PE production is very strongly suppressed by the combined effects of the Rca system, which operates via transcriptional repression, and the Cgi system, which also represses production, but does so posttranscriptionally. Conversely, in the CA2 species N. punctiforme, only a single system regulates PE expression, which is never completely repressed in red light (8, 18). This initial study also noted significant differences in the amount of PE and PC produced in different species, and within both the CA2 and CA3 groups (8). This complexity in CA2 and CA3 responses is likely due to promoter strength differences as well as the types and numbers of CA regulatory pathways used by each cyanobacterial species.

### Materials and Methods

#### Growth Conditions.

Growth conditions. SF33 (54) of Fremyella diplosiphon UTEX 481 (also called Tolypothrix sp. PCC 7601) was wild type. Cultures were grown as described (15) or with 10 μg/mL kanamycin in 15 μl photons m⁻² s⁻¹ using red and green fluorescent lights (Industrial F40T12-Red and -Green, Light Bulbs Unlimited).

#### RNA Analysis.

For RNA half-life measurements, rcaE cells were grown to midlogarithmic phase in red or green light. Rifampicin was dissolved in DMSO at 200 mg/mL and added to cultures at a final concentration of 1 μg/mL in the same light after rifampicin addition. Samples (50 mL) were taken immediately after rifampicin addition (time zero) and after 1, 2, 5, 10, 20, 40, and 60 min. RNA was isolated and analyzed as described (15) except that 0.9 mL of Tris- Reagent was used during isolation and RNA was precipitated by adding a 0.5× vol isopropanol and loading onto a Qiagen RNaseq Mini kIt column, then eluted per the manufacturer's instructions. The same method, minus rifampicin, was used to isolate RNA from wild-type, rcaE, and rcaC cultures grown in red and green light. RNA was separated either by electrophoresing 15 μg of each sample for 2 h at 100 V on 1% agarose-formaldehyde gels (55) then transferring overnight to Immobilon Nytran-NY+ membrane (Fisher Scientific) using 10x SSC or by resuspending 5 μg of each sample in 10 μl of Sample Buffer (20 mM Mops, pH 7.01 mM di-sodium EDTA5 mM sodium acetate50% formamide/70% formamide/0.040 μg/mL ethidium bromide) plus 2 μl of Sample Dye (50% glycerol/10 mM di-sodium EDTA/2.5 mg/mL bromophenol blue/2.5 mg/mL xylene cyanol), heating to 65 °C for 5 min and electrophoresing for 1.5 h at 100 V on a 6%, 8 μM urea-Tris-borate-EDTA polyacrylamide gel. RNA was then transferred to Immobilon Nytran-NY+ at 15 V for 1 h (55). After UV cross-linking, membranes were probed as previously described. PCR amplification of F. diplosiphon genomic DNA using primers cpeC-L and cpeC-R (primers shown in Table 5) generated the cpeC coding sequence and primers cpeC′L and cpeC′R were used to amplify a cpeC 5′ leader probe (15). A Molecular Dynamics SP Phosphomager was used to quantify probe hybridization and ribosomal values were used to normalize mRNA values.

#### Plasmid Construction.

All numbering is relative to the transcription start site of the gene involved. pRB7 was made by PCR amplification using primers 400cpeC and CDEpcboxmut1 in one PCR amplification and cpeCUTR and CDEpcboxmut2 in another. The two resulting PCR products were annealed and PCR amplification using 400cpeC and cpeCUTR, and the product cut with Sph and BamHI and inserted into similarly cut pRB1. Two-step PCR amplification was also used to create pRB13 (primer pairs 400cpeC/cpeCstem1-1 and cpeCUTR/cpeCstem1-2). pRB14 was also used to create pRB13 (primer pairs 400cpeC/cpeCstem1-1 and cpeCUTR/cpeCstem1-2). pRB15 (primer pairs 400cpeC/cpeCstem3-1 and cpeCUTR/cpeCstem3-2). pRB12 was synthesized using primer pairs 400cpeC/cpeCstem4-1 in a PCR amplification. The product was used as a template for a second amplification, using primer pairs 400cpeC/cpeCstem4-2. The final PCR amplification products for pRB12-pRB15 were cut and ligated into pRB1 as described for pRB7 and pRB8. Stem-loop 4 mutations were made by PCR amplification using pRB8 and primer 400pexc paired with one of the following four primers: Stem4onlmuty, Strstem4, 4Loopmut, and Allp4air. Products were used in a second PCR amplification with primer pair 400cpeC/cpeCstem4-2. The resulting fragments were cut with Sph and BamHI and cloned into the same sites in pRB1. Sequences 3′ of stem-loop 4 were changed using pRB8 and PCR amplifying with primer pair 400cpeC/Down4mut2. The first amplification product was reamplified with primer pair 400cpeC/ Down4mut2. The product was cut with Sph and BamHI and cloned into the same sites in pRB8. Plasmid transformations into F. diplosiphon were by conjugation (16, 54). All PCR-amplified DNA and ligation junctions were sequenced.

#### GUS Assays.

GUS assays were modified from previous protocols (29, 31). Transformants were grown in BG-11 with kanamycin to an A750 of 0.7 in either red or green light. A total of 350 μL was centrifuged for 4 min at 16,000 × g at room temperature. Pellets were resuspended in 1 mL of GUS assay buffer (50 mM NaPO₄, pH 7.01 mM EDTA) containing 12 μg/mL chlorophenol and centrifuged as before. Pellets were resuspended in 1 mL of GUS assay buffer, 20 μl of 0.1% SDS, and 40 μl of chloroform, then sequenced.
vortexed for 10 s. At least three technical replicates were conducted for each transformant, and each construct was assayed using at least four independently transformed lines. For each replicate, 20 µl of cell lysate was mixed with 180 µl of GUS assay buffer containing 1.25 mM X-glutathionyl 1-glucuronyl (PNGP, Sigma) and incubated at room temperature. Absorbance was measured at 405 nm and recorded every 2 min for 30 min with a Molecular Devices SpectraMax 190 (Molecular Dynamics). Protein concentrations determined using a Pierce BCA protein assay reagent kit per the manufacturer’s instructions. Activity was quantified as nmol of product per mg of protein per min.

3' RACE Analysis. Methods for 3' RACE analysis are described in SI Materials and Methods.

cpeC 5’ Leader ORF Disruption. Methods for cpeC 5’ leader ORF disruption are described in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank the D.M.K. laboratory members for thoughtful discussions and comments on the manuscript. This work was supported by National Science Foundation Grant MCB-1029414 (to D.M.K.).

38. Fargo DC, Boynton JE, Gillham NW (1999) Mutations altering the predicted secondary structure of a chloroplast 5' untranslated region affect its physical and biochemical properties as well as its ability to promote translation of reporter mRNAs both in the Chlamydomonas reinhardtii chloroplast and in Escherichia coli. Mol Cell Biol 19:6890–6890.
**Supporting Information**

**Bezy et al. 10.1073/pnas.1107427108**

**SI Materials and Methods**

**3′ RACE Analysis.** RNA from green-light grown rcaC− (CR2) cells was extracted as described in Methods and Materials except a QIAGEN miRNeasy kit was used. Poly(A) tails were added to the 3′ ends of the mRNA using the mRNA-ONLY Prokaryotic mRNA Isolation Kit with Poly(A)-Tailing (Epicentre) per manufacturer’s instructions using 2 μg of total RNA. Terminator exonuclease and poly(A) polymerase reactions were stopped by adding a 1:1:1 ratio of reaction:chloroform:phenol. Products were precipitated with 1.5x vol 100% ethanol at −20 °C for 30 min and isolated on an miRNasy column. The RNA was reverse transcribed for 1 h at 37 °C using the ExactSTART Eukaryotic mRNA 5′- and 3′-RACE kit from Epicentre. After RNase treatment, PCR amplification was performed using the RACEcpeCfor primer, specific for the 5′ end of the cpeC transcript (5′−AACAAAGGTATAAAGCCTAAAACCTT−3′) and PCR primer 2 from the ExactSTART kit. Forty cycles were run as: 98 °C for 30 s; 57 °C for 30 s; 72 °C for 30 s. Product was electrophoresed on a 1% agarose gel. The single predominant band from the primer 2 reaction containing cDNA was gel extracted and sequenced using RACE-cpeCfor as the primer.

**cpeC 5′ Leader ORF Disruption.** Constructs used to test the role of the ORFs within the cpeC 5′ leader in Cgi system function were made using pRB8 and the same approach as outlined in the Plasmid Construction section of Materials and Methods. Stop codons were introduced using PCR amplification, one using primer 400cpeC paired with primer cpeCorfX1, cpeCorfY1, or cpeCorfZ1, and the other using primer cpeCutr2 paired with either cpeCorfX2, cpeCorfY2, or cpeCorfZ2. Primers changed the second or third codon of each ORF to a stop codon. Primer-generated SphI and BamHI sites were cut and used to insert products into similarly cut pRB8.

---

**Fig. S1.** A single transcript encompasses the entire.cpeCDESTR operon. (A) Representative autoradiographs of RNA blot analyses using probes hybridizing to the coding region of cpeC (Left) or cpeR (Right). Similar results were obtained for three independent biological replicates. Corresponding ribosomal bands are shown (ribo). (B) Diagrammatic representation of the cpeC operon and major transcripts detected by RNA blot hybridization analyses in A. The thickness of each arrow indicates its abundance relative to the other transcripts produced. Bent arrow indicates the transcription start site.
Fig. S2. The effect of stem4 position on Cgi regulation of cpeC. (A) Diagrammatic representation of base construct (pRB8) used to make changes in the position of stem4. PCR amplifications were conducted as described in Materials and Methods using the four primers listed in Table S1. Bent arrow indicates transcription start site. (B) Sequence changes made in the region containing stem4 of the cpeC 5′ leader that shifted stem4 −10, −5, +5, and +10 relative to its original position. Underlined and bolded sequences indicate the stem4 region; boxed areas denote the putative Shine Dalgaro (SD) and translation start codon. (C) Relative mean rates of GUS activity from lysates of F. diplosiphon transformed with the indicated plasmids and grown in red light (RL) or green light (GL). The mean value (255.9 nmol of product per mg of protein per min) derived from cells transformed with pRB8 and grown in green light was set at 100%. At least four independently transformed lines were tested for each plasmid and light condition except for pRB8, which was tested three times. Error bars show SE.

Fig. S3. Analysis of the possible role of ORFs in the 5′ leader of cpeC on Cgi system regulation of GUS activity. (A) Diagrammatic representation of base construct (pRB8) used to introduce a stop codon into the three major ORFs (small arrows) within the coding strand of the cpeC 5′ leader. PCR amplifications were conducted as described in Materials and Methods using the four primers listed in Table S1. The ORFs encompassed the following regions relative to transcription start: ORFX, +55 to +120; ORFY, +107 to +172; ORFZ, +117 to +161. Bent arrow indicates transcription start site. (B) Relative mean rates of GUS activity from lysates of F. diplosiphon transformed with the indicated plasmids and grown in red light (RL) or green light (GL). The mean value (242 nmol of product per mg of protein per min) derived from cells transformed with pRB8 and grown in green light was set at 100%. At least four independently transformed lines were tested for each plasmid and light condition. Error bars show SE.
Fig. S4. 3′ RACE of RNA from rcaC mutant cells to check for the presence of RNAs shorter than cpeCD. (A) Ethidium bromide stained agarose gel containing 3′ RACE products using the primer indicated. Negative control is to the right. Arrow indicates the band excised and used for sequencing. (B) Diagrammatic representation of the 413 bp region of the cpeC gene identified as the 3′ RACE product, sequenced using the DNA from the gel band excised in A. Bent arrow, transcription start; black box, stem4 region.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Construct</th>
<th>Sequence (5′ to 3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cpeC-L</td>
<td>AGCAGAAAAAATCCATTCGAGCGCA</td>
<td></td>
</tr>
<tr>
<td>cpeC-R</td>
<td>TGCTTGGTGTGCTTTGAAATC</td>
<td></td>
</tr>
<tr>
<td>cpeC5′ L</td>
<td>AAAAAAGCCTTAAAAACGCTTAAAA</td>
<td></td>
</tr>
<tr>
<td>cpeC5′ R</td>
<td>CAGTTGTTAAATAATTTACGATTGCT</td>
<td></td>
</tr>
<tr>
<td>400cpeC</td>
<td>gcggcgcCTAATGCCTGACACATCGTGAGGT (Sphi site underlined)</td>
<td></td>
</tr>
<tr>
<td>cpeCUTR</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
<td></td>
</tr>
<tr>
<td>CDEpcboxmut1</td>
<td>pR88</td>
<td>AATTGAGAAGAATGTAAGCTCAATGTAAGCAAAACTATTTGAA (mutated L Box in bold)</td>
</tr>
<tr>
<td>CDEpcboxmut2</td>
<td>pR88</td>
<td>TAAACTTACTACGTTGACTACTACACATGTAATTGCTTAAGAAAAGCTTTAGTTGATGCTTCAAGGCTTTCATCA</td>
</tr>
<tr>
<td>cpeCstem1-1</td>
<td>pR13</td>
<td>TTGTTCGATTGTTAAATTTTACGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>cpeCstem1-2</td>
<td>pR13</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>cpeCstem2-1</td>
<td>pR14</td>
<td>CTAAAACCTTGGCATTCAGGAAATCAGTATTTTGTTGCAACAAATCAATCC</td>
</tr>
<tr>
<td>cpeCstem2-2</td>
<td>pR14</td>
<td>TGGTTTTCTTGCAAGTTTTTACGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>cpeCstem3-1</td>
<td>pR15</td>
<td>CTATTGCTGTTGTTAAGGCTGTTAAGATGACACACGTTGACTTTATTTCCCATCC</td>
</tr>
<tr>
<td>cpeCstem3-2</td>
<td>pR15</td>
<td>TACCGTTGCTATAGCAGTAAAGACGTTGACTTTATTTCCCATCC</td>
</tr>
<tr>
<td>cpeCstem4-1</td>
<td>pR12</td>
<td>AATTGAGAAGAATGTAAGCTCAATGTAAGCAAAACTATTTGAA (mutated L Box in bold)</td>
</tr>
<tr>
<td>cpeCstem4-2</td>
<td>pR12</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>Stem4onlymut</td>
<td>pR21</td>
<td>GTTTCTTGCGTAAAAGAAATTTTATTTCGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>StrSt4</td>
<td>pR19</td>
<td>GTTTCTTGCGTAAAAGAAATTTTATTTCGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>Allpair4</td>
<td>pR20</td>
<td>GTTTCTTGCGTAAAAGAAATTTTATTTCGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>4LoopMut</td>
<td>pR18</td>
<td>GTTTCTTGCGTAAAAGAAATTTTATTTCGATTGCTTTAGTTGAA (BamHI site underlined)</td>
</tr>
<tr>
<td>Down4mut1</td>
<td>pR22</td>
<td>TACCGTTGCTATAGCAGTAAAGACGTTGACTTTATTTCCCATCC</td>
</tr>
<tr>
<td>Down4mut2</td>
<td>pR22</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>−5St4shift</td>
<td>pR24</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>−10St4shift</td>
<td>pR23</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>+5St4shift</td>
<td>pR25</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>+10St4shift</td>
<td>pR26</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
</tr>
<tr>
<td>cpeCorfX1</td>
<td>orfX</td>
<td>CAAATTACAGCTACATAAGCATGAAA</td>
</tr>
<tr>
<td>cpeCorfY1</td>
<td>orfY</td>
<td>TGCTACATGAGTCATGACATTCT</td>
</tr>
<tr>
<td>cpeCorfZ1</td>
<td>orfZ</td>
<td>AATAAATTCGGATTAGCTCCAGGACG</td>
</tr>
<tr>
<td>cpeCutr2</td>
<td>gcggcgciiCCAAATGTCATTAAATTCCCTGCT (BamHI site underlined)</td>
<td></td>
</tr>
<tr>
<td>cpeCorfX2</td>
<td>orfX</td>
<td>TTTAGCATTGAAAGCTGCTGTTATTGT</td>
</tr>
<tr>
<td>cpeCorfY2</td>
<td>orfY</td>
<td>AAGAATGTAGTGACTGAGTAGGA</td>
</tr>
<tr>
<td>cpeCorfZ2</td>
<td>orfZ</td>
<td>GCCTGAAGACTCTCAATGGAAATTTT</td>
</tr>
</tbody>
</table>

Lowercase letters denote added sequences.