Role of the G-box element in regulation of chlorophyll biosynthesis in arabidopsis roots

Koichi Kobayashi,^{1,*} Takeshi Obayashi² and Tatsuru Masuda¹

¹Graduate School of Arts and Sciences; The University of Tokyo; Tokyo, Japan; ²Graduate School of Information Sciences; Tohoku University; Sendai, Japan

Keywords: chlorophyll biosynthesis, photosynthesis, transcriptional regulation, cis-element, co-expression network

Regulation of chlorophyll metabolism comprises strong transcriptional control together with a range of post-translational mechanisms during chloroplast biogenesis. Recently we reported that chlorophyll biosynthesis in *Arabidopsis thaliana* roots is regulated by auxin/cytokinin signaling via the combination of two transcription factors, LONG-HYPOCOTYL5 (HY5) and GOLDEN2-LIKE2 (GLK2). In this study, we examined the involvement of cis-elements in the expression of chlorophyll biosynthesis genes. Searches for predicted cis-elements in key chlorophyll biosynthesis genes and their co-expressed genes revealed coexistence of the G-box motif and the CCAATC motif, which may be targeted by HY5 and GLK factors, respectively, in their promoter regions. Deletion of the G-box from the promoter of the *CHLH* gene encoding the H subunit of Mg-chelatase resulted in the absence of its expression in roots but not in shoots, showing a differing involvement of the G-box in *CHLH* expression between shoots and roots. Our data suggest that transcription factors and cis-elements participating chlorophyll biosynthesis are substantially changed during organ differentiation, which may be linked to the differentiation of plastids.

It is widely acknowledged that plant plastids evolved from an ancestral cyanobacterium that was engulfed by a protoeukaryotic cell containing mitochondria.¹ Plastids in unicellular algae usually function as chloroplasts and harvest energy from sunlight by photosynthesis. On the other hand, plastids in higher plants undergo profound morphological changes to generate different plastid types. Whereas chloroplasts prevail in leaves, non-green plastids (e.g., amyloplasts, chromoplasts) develop in non-photosynthetic organs such as roots and flowers. Thus, higher plants have developed regulatory systems to coordinate plastid differentiation with cell differentiation.¹

At the onset of chloroplast biogenesis, chlorophylls and photosynthetic proteins are actively synthesized, with the concomitant development of thylakoid membrane networks. Although chlorophylls are essential for light energy harvesting and charge separation in photosystems, chlorophylls and their intermediates are strong photosensitizers that produce highly reactive singlet oxygen when they are excited to a triplet state by light.² Therefore, plant cells strictly regulate tetrapyrrole metabolism, including chlorophyll biosynthesis, to avoid excessive accumulation of these pigments and consequent photooxidative damage. Regulation of chlorophyll biosynthesis includes post-translational modifications of enzyme activities in chloroplasts and transcriptional regulation in the nucleus.³ Whereas post-translational regulation mainly occurs at particular metabolic steps such as the biosynthesis of 5-aminolevulinic acid (the first committed step of tetrapyrrole biosynthesis) or the insertion of Mg²⁺ into protoporphyrin IX (the first step of the chlorophyll

biosynthetic branch), transcriptional regulation appears to occur coordinately throughout the tetrapyrrole metabolic pathway. In fact, it is reported that the expressions of chlorophyll biosynthetic genes, which are all encoded in the nucleus in angiosperms, are globally downregulated when the chloroplast function is severely impaired.⁴ Analyses of gene expression networks revealed that *HEMA1* (encoding glutamyl-tRNA reductase), *CHLH* (encoding the H subunit of Mg-chelatase), *GENOMES UNCOUPLED4* (*GUN4*), *CHL27* (encoding the membrane subunit of Mg-protoporphyrin IX monomethyl ester cyclase), *CHLOROPHYLL A OXYGENASE* (*CAO*) and *CHLP* (encoding geranylgeranyl pyrophosphate reductase) are particularly highly co-expressed with nuclear-encoded photosynthesis-related genes in response to environmental and developmental stimuli such as light and circadian rhythms.^{5,6} Thus, it is proposed that this regulatory co-expression plays a central role in the assembly of the photosynthetic machinery during chloroplast biogenesis.

Recently, we reported that chlorophyll biosynthesis in *Arabidopsis* roots is regulated by auxin/cytokinin signaling at the transcriptional level via the combined action of two transcription factors, LONG-HYPOCOTYL5 (HY5) and GOLDEN2-LIKE2 (GLK2).⁷ HY5 is required for the expression of key chlorophyll biosynthetic genes in roots whereas GLK2 strongly upregulates these genes. Moreover, characterization of *hy5* mutants overexpressing *GLK1* or *GLK2* revealed that GLK factors require HY5 to maximize their function in the root. Cis-element analyses showed that HY5 binds to the G-box (CACGTG) in the promoter regions of target genes⁸ whereas GLK factors recognize

*Correspondence to: Koichi Kobayashi; Email: kkobayashi@bio.c.u-tokyo.ac.jp
Submitted: 04/12/12; Revised: 05/14/12; Accepted: 05/15/12
<http://dx.doi.org/10.4161/psb.20760>

Figure 1. (A) Expression networks of the nuclear-encoded genes that are highly co-expressed with *CHLH*, *CHL27* and *CHLP*. The networks were drawn using the ATTED-II database. (B) DNA motifs extracted from the promoter sequences of the 20 co-expressed genes. The size of the letter reflects the frequency of the corresponding base at each position. (C) Structure of *CHLH promoter::GUS* fusion genes with or without the G-box in the promoter. Black and gray lines indicate 5'-upstream regions and untranslated regions, respectively, whereas open boxes show the regions of the first three amino acids derived from the *CHLH* gene. GUS coding regions are depicted in blue. The nucleotides are numbered relative to the transcription start site in wild type *CHLH*. In the *CHLH_{pro}ΔG::GUS* line, the G-box composed of CACGTG was deleted from the promoter region. (D) Histochemical analysis of GUS activity in 21-d-old transgenic plants carrying each construct.

a putative CCAATC motif.⁹ Considering that protein-protein interactions were observed between GLK proteins and G-box binding factors,¹⁰ the combination of HY5 on the G-box and GLK factors on the CCAATC motif may be important for the coordinated regulation of chlorophyll biosynthetic and photosynthetic genes. However, little is known about mode of action of these transcription factors on their target promoters.

To gain insight into the mechanism for coordination of nuclear-encoded photosynthesis-related gene expression, we performed in silico predictions of cis-elements that could be involved in the co-expression networks. As previously reported,⁵ strong correlation of expression among *CHLH*, *CHL27* and *CHLP* (MR = 1.4–2.5, PCC = 0.91–0.92) was found using the “EdgeAnnotation” tool in a co-expression database for Arabidopsis, ATTED-II.^{11,12} In

addition, 17 genes that were strongly co-expressed with these three genes were identified using the “CoExSearch” tool. Most of these 17 genes were related to photosynthesis and formed tight co-expression networks with each other (Fig. 1A). From the proximal promoter sequences around the transcription start site from -500 to +300 of the 20 genes, the Multiple Em for Motif Elicitation (MEME) software¹³ extracted 3 motifs, two of which were similar to motifs generally called CT-repeats and one of which was similar to G-boxes (Fig. 1B).

If these motifs function as cis-elements to regulate the 20 genes in common, it is expected that the cis-elements would not randomly appear elsewhere in the genome. Therefore, we investigated the specificity of the three motifs in terms of three aspects. The first aspect was the specificity of these candidate cis-elements to the 20 genes of interest among all genes in Arabidopsis. We searched for genes containing the candidate cis-elements using the Motif Alignment and Search Tool (MAST) software.¹³ The second aspect was the existence of position preference, because most reported cis-elements have position preferences within the promoter sequence.¹⁴ The third aspect was the gene functions associated with the motifs, which were evaluated using Gene Ontology (GO) annotations in The Arabidopsis Information Resource (TAIR).¹⁵ If these motifs were functional cis-elements, some functional bias should be observed in the genes containing the motifs. The results showed that specificities of the CT-repeats were high for position, but low for the other two parameters, whereas the G-box possessed high specificities in all three parameters (Table 1). Our data suggest an involvement of the G-box in the co-expression system for nuclear photosynthetic genes.

In addition to the MEME analysis, we also applied the k-tuple method to directly compare the promoter sequences of the 20 co-expressed photosynthesis genes with those of the other genes in the Arabidopsis genome. Table 2 shows the results of 6-, 7- and 8-tuples significantly accumulated in the 20 co-expressed gene promoters. Interestingly, not only the G-box but also the putative GLK recognition motif (CCAATC) was identified as a potential cis-element of the co-expressed genes. Moreover, these two motifs coexist in the promoter regions of the co-expressed genes with significant frequency (Table 3), implying that these cis-elements function together. This idea is consistent with the report that protein-protein interactions occur between GLK proteins and G-box binding factors.

A chromatin immunoprecipitation chip analysis identified genome-wide in vivo HY5 binding targets, which tend to be enriched in early light-responsive genes including photosynthetic genes and transcription factor genes.⁸ In fact, most nuclear genes encoding subunits of photosynthetic complexes are putative HY5 targets. In addition, among the six key chlorophyll biosynthesis genes (*HEMA1*, *CHLH*, *GUN4*, *CHL27*, *CAO* and *CHLP*) involved in the co-expression network,⁵ five genes were identified

Table 1. Prediction of cis-elements in the promoter regions of co-expressed photosynthetic genes

Motif	Query number ^a	E of motif ^b	GHM in query ^c	GHM in all genes ^d	p for motif enrichment ^e	p for position preference ^f	P for GO test ^g	GO annotation ^h
Motif 1 (CT-repeat)	20	5.00E-11	1	467	0.28	0	0.10	translational elongation
Motif 2 (G-box)	13	2.70E-09	4	86	8.91E-07	0.01	8.32E-05	thylakoid
Motif 3 (CT-repeat)	20	4.20E-09	1	267	0.19	8.91E-16	1.00	

^aThe number of query genes used for motif extraction by the MEME program. ^bThe e-value of the motifs extracted by the MEME program. ^cThe number of genes having the predicted motifs (GHM) in their promoter regions among the query genes (MAST score > 1,000). ^dThe number of GHM in their promoter regions among all Arabidopsis genes (MAST score > 1,000). ^eThe p-value of a hypergeometric test for motif enrichment on the query genes against all Arabidopsis genes, which reflects the specificity of these motifs in the query genes. ^fThe p-value of Kolmogorov-Smirnov test for a position preference of these motifs in the promoter region. ^gThe most significant p-value for GO term enrichment tests (hypergeometric test) with Bonferroni multi-test correction for the number of GO terms tested. ^hThe most significant GO annotation related to the respective motif.

as HY5-targeted genes, with *HEMA1* as the only exception. Moreover, our recent analysis revealed that the expressions of *CHLH*, *CHL27* and *CHLP*, but not *HEMA1*, were substantially reduced in roots of the *hy5* mutant.⁷ This suggested that the binding of HY5 to G-box elements is one of the central mechanisms for coordinating the transcription of key genes involved in chlorophyll biosynthesis and photosynthesis, at least in the root.

To test this hypothesis, we chose *CHLH* as a representative gene for the analysis and constructed *CHLH* promoter-*GUS* reporter lines of Arabidopsis containing wild type (*CHLH_{pro}*) or G-box-lacking promoters (*CHLH_{pro}ΔG*) (Fig. 1C). As shown in Figure 1D, intense GUS staining was observed in the roots of the *CHLH_{pro}:GUS* line, while almost no staining was detected in the roots of the *CHLH_{pro}ΔG:GUS* line. These staining patterns were consistently observed in several transgenic lines for each construct. These results clearly demonstrate that the G-box element, presumably binding HY5, is essential for *CHLH* expression in the root. This is probably applicable to other co-expressed genes as well, although direct binding of HY5 to the G-box elements of these gene promoters, including the *CHLH* promoter, has not been confirmed yet. It is interesting to note that in aerial organs, the levels of GUS staining were not markedly different between these two lines, suggesting a lesser importance of G-box binding factors in the *CHLH* expression in the shoot. The data are in good agreement with the fact that chlorophyll accumulation in the leaves is not strongly impaired even in the double mutant for *HY5* and *HY5 HOMOLOG*,¹⁶ whereas that in *hy5* roots is almost negligible.^{7,17,18} By contrast, the chlorophyll content in the *glk1 glk2* double mutant is reduced by 50% in roots⁷ but by 80% in leaves¹⁹ as compared with wild type, showing that these factors play a more significant role in the shoot. Meanwhile, overexpression of GLKs substantially increased chlorophyll content in the root only in the presence of *HY5*,⁷ suggesting that these factors have potential to induce chloroplast biogenesis in non-photosynthetic tissues. Although the role of chlorophyll biosynthesis and chloroplast biogenesis in the root remains unclear, it is possible that HY5 and GLK factors regulate chloroplast biogenesis plastically in response to environmental conditions such as light availability and energy demand, particularly when the aerial photosynthetic tissues suffer serious damage.

In conclusion, our data indicate that transcription factors as well as cis-elements participating in the regulation of chlorophyll biosynthesis are vastly different between the shoot and the root. Considering that the G-box and the putative GLK recognition

Table 2. Sequences enriched in the 20 co-expressed gene promoters

Tuples	p-value	Related element
CCACGTG	1.2E-02	G-box
CCACGT	3.4E-05	
GCCACGT	3.0E-03	
GCCACG	1.0E-02	
ACCAATCA	6.7E-03	GLK recognition
CCAATC	4.9E-02	
ACCAATC	9.5E-04	
ACCAAT	3.8E-02	
CTTATCCA	4.9E-02	unknown
TATCCA	6.5E-04	
AATGGC	3.8E-02	unknown
AAATCC	2.9E-02	unknown

element coexist in the promoter regions of the co-expressed genes with high frequency (Table 3), it is possible that the importance and/or mode of interactions of transcription factors on these two cis-elements vary between the shoot and the root. Changes in sets of transcription factors and cis-elements involved in chlorophyll biosynthesis between photosynthetic and non-photosynthetic organs may be associated with the coordinated regulation of plastid differentiation and cell differentiation.

GUS Assay

The 5'-upstream region of *CHLH* between -1,200 and +155 bp from the transcription initiation site (*CHLH_{pro}*) was amplified from the Arabidopsis Col-0 genome by PCR using specific primers (5'-AAG CTT CCA GGT TAA GCT ATA CGG AC-3' and 5'-GGA TCC CGA AGC CAT TTT GCG GC-3') and cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's instructions. To construct the *CHLH* promoter lacking G-box sequences (*CHLH_{pro}ΔG*), inverse PCR was performed using the circular *CHLH_{pro}*-pGEMT-T Easy plasmid as the template and primers (5'-TCC TTC CCT CTC ACC ACT TAC C-3' and 5'-GAC GAA TAG TTA CGG AAG AAG GGA G-3') which were designed to hybridize in a tail-to-tail fashion at locations adjacent to the G-box sequences in the *CHLH* promoter region. The resulting PCR product lacking the G-box was self-ligated into a circular plasmid after phosphorylation by a T4 polynucleotide kinase (TaKaRa). After checking the sequences, these fragments were ligated into the *GUS* gene in the HindIII/

Table 3. Coexistence of G-box and GLK recognition motifs in the co-expressed gene promoters

Motif	GHM/co-expressed genes ^a	GHM/all genes ^b	p-value ^c
CCACGT (G-box)	12/20	3351/31374	8.6E-09
CCAATC (GLK recognition)	12/20	6206/31374	1.3E-05
CCACGT and CCAATC	10/20	701/31374	9.0E-14

^aThe number of genes having the motifs (GHM) in their promoter regions among the 20 co-expressed genes. ^bThe number of GHM in their promoter regions among all Arabidopsis genes. ^cp-value for motif enrichment analysis using hypergeometric test.

BamHI site of the pBI101 vector. T-DNA regions carrying these constructs were introduced into the Arabidopsis Col-0 genome using *Agrobacterium tumefaciens* strain GV3101::pMP90.

Transgenic Arabidopsis plants were grown vertically on solid medium (1x Murashige and Skoog [MS] medium, 1% [w/v] sucrose and 0.8% [w/v] agar, pH 5.7) for 21 d at 23°C under continuous light. Histochemical analysis for GUS expression was performed as previously described in reference 20, at least in nine independent transgenic lines for each construct.

Cis-element Prediction

The co-expression analysis using ATTED-II (<http://atted.jp>) was described by Obayashi et al.^{11,12} The genome sequences around the transcription start site from -500 to +300 of target

genes were downloaded from the MIPS database²¹ and then cis-element motifs were extracted by the MEME software.¹³ Specificity and functional bias of the cis-element candidates were analyzed using MAST software¹³ and GO annotations downloaded from the TAIR website,¹⁵ respectively. All GO annotations were first mapped to all upper GO terms up to the root terms. Hypergeometric tests were applied to test the two aspects. Positional distributions of the cis-element candidates were tested using the Kolmogorov-Smirnov test against uniform distribution. As another cis-element prediction method, we also applied the k-tuple method to directly compare the promoter sequences of the 20 photosynthesis genes of interest with those of the other genes in the genome. The 6-, 7- and 8-tuple was used in this study. After counting each tuple in the photosynthetic gene promoters and in the other gene promoters, hypergeometric test with bonferroni correction was applied to select significantly accumulated tuples in the co-expressed photosynthetic gene promoters.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgements

Support by Grants-in-Aid for Scientific Research (No. 21570035 to T.M.) and in part by the Global Center of Excellence Program (K03) from the Ministry of Education, Culture, Sports and Technology (MEXT), Japan, and a Research Fellowship for Young Scientists from the Japanese Society for the Promotion of Science (JSPS) and a RIKEN postdoctoral fellowship to K.K. are acknowledged.

References

- Lopez-Juez E, Pyke KA. Plastids unleashed: their development and their integration in plant development. *Int J Dev Biol* 2005; 49:557-77; PMID:16096965; <http://dx.doi.org/10.1387/ijdb.051997el>.
- Triantaphylidès C, Havaux M. Singlet oxygen in plants: production, detoxification and signaling. *Trends Plant Sci* 2009; 14:219-28; PMID:19303348; <http://dx.doi.org/10.1016/j.tplants.2009.01.008>.
- Tanaka R, Kobayashi K, Masuda T. Tetrapyrrole Metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* 2011; 9:145; PMID:22303270; <http://dx.doi.org/10.1199/tab.0145>.
- Moulin M, McCormac AC, Terry MJ, Smith AG. Tetrapyrrole profiling in Arabidopsis seedlings reveals that retrograde plastid nuclear signaling is not due to Mg-protoporphyrin IX accumulation. *Proc Natl Acad Sci USA* 2008; 105:15178-83; PMID:18818314; <http://dx.doi.org/10.1073/pnas.0803054105>.
- Masuda T, Fujita Y. Regulation and evolution of chlorophyll metabolism. *Photochem Photobiol Sci* 2008; 7:1131-49; PMID:18846277; <http://dx.doi.org/10.1039/b807210h>.
- Matsumoto F, Obayashi T, Sasaki-Sekimoto Y, Ohta H, Takamiya K, Masuda T. Gene expression profiling of the tetrapyrrole metabolic pathway in Arabidopsis with a mini-array system. *Plant Physiol* 2004; 135:2379-91; PMID:15326282; <http://dx.doi.org/10.1104/pp.104.042408>.
- Kobayashi K, Baba S, Obayashi T, Sato M, Toyooka K, Keränen M, et al. Regulation of root greening by light and auxin/cytokinin signaling in Arabidopsis. *Plant Cell* 2012; 24:1081-95; PMID:22415275; <http://dx.doi.org/10.1105/tpc.111.092254>.
- Lee J, He K, Stolz V, Lee H, Figueroa P, Gao Y, et al. Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 2007; 19:731-49; PMID:17337630; <http://dx.doi.org/10.1105/tpc.106.047688>.
- Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *Plant Cell* 2009; 21:1109-28; PMID:19376934; <http://dx.doi.org/10.1105/tpc.108.065250>.
- Tamai H, Iwabuchi M, Meshi T. Arabidopsis GARP transcriptional activators interact with the Pro-rich activation domain shared by G-box-binding bZIP factors. *Plant Cell Physiol* 2002; 43:99-107; PMID:11828027; <http://dx.doi.org/10.1093/pcp/pcf011>.
- Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K. ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res* 2009; 37:987-91; PMID:18953027; <http://dx.doi.org/10.1093/nar/gkn807>.
- Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, et al. ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res* 2007; 35:863-9; PMID:17130150; <http://dx.doi.org/10.1093/nar/gkl783>.
- Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 2006; 34:369-73; PMID:16845028; <http://dx.doi.org/10.1093/nar/gkl198>.
- Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, et al. Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 2007; 8:67; PMID:17346352; <http://dx.doi.org/10.1186/1471-2164-8-67>.
- Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* 2008; 36:1009-14; PMID:17986450; <http://dx.doi.org/10.1093/nar/gkm965>.
- Sibout R, Sukumar P, Hettiarachchi C, Holm M, Muday GK, Hardtke CS. Opposite root growth phenotypes of hy5 versus hy5 hyh mutants correlate with increased constitutive auxin signaling. *PLoS Genet* 2006; 2:202; PMID:17121469; <http://dx.doi.org/10.1371/journal.pgen.0020202>.

17. Oyama T, Shimura Y, Okada K. The *Arabidopsis* HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev* 1997; 11:2983-95; PMID:9367981; <http://dx.doi.org/10.1101/gad.11.22.2983>.
18. Usami T, Mochizuki N, Kondo M, Nishimura M, Nagatani A. Cryptochromes and phytochromes synergistically regulate *Arabidopsis* root greening under blue light. *Plant Cell Physiol* 2004; 45:1798-808; PMID:15653798; <http://dx.doi.org/10.1093/pcp/pch205>.
19. Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA. GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J* 2002; 31:713-27; PMID:12220263; <http://dx.doi.org/10.1046/j.1365-313X.2002.01390.x>.
20. Kobayashi K, Awai K, Takamiya K, Ohta H. *Arabidopsis* type B monogalactosyldiacylglycerol synthase genes are expressed during pollen tube growth and induced by phosphate starvation. *Plant Physiol* 2004; 134:640-8; PMID:14730084; <http://dx.doi.org/10.1104/pp.103.032656>.
21. Schoof H, Ernst R, Nazarov V, Pfeifer L, Mewes HW, Mayer KF. MIPS *Arabidopsis thaliana* Database (MAtdB): an integrated biological knowledge resource for plant genomics. *Nucleic Acids Res* 2004; 32:373-6; PMID:14681437; <http://dx.doi.org/10.1093/nar/gkh068>.

© 2012 Landes Bioscience.
Do not distribute.