Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in Arabidopsis

Me-Jeong Jeong and Ming-Che Shih *

Department of Biological Sciences, 204 Chemistry Building, University of Iowa, Iowa City, IA 52242, USA

Received 22 November 2002

Abstract

We have previously identified a cis-acting element, named the XXIII box, that is essential for light-regulated expression of the nuclear gene GAPB, which encodes the B subunit of chloroplast glyceraldehyde-3-phosphate dehydrogenase from Arabidopsis thaliana. Examination of the sequences indicated that there are two GATA motifs within the XXIII box. Based on the degree of the amino-acid sequence identity in the DNA binding domains, we divided the 25 GATA factors encoded in the Arabidopsis genome into three classes. We chose GATA-1 and GATA-20 from Class I and Class II, which include the majority of GATA factors, for overexpression in an Escherichia coli expression system. Gel mobility shift assays showed that GATA-1, but not GATA-20, binds specifically to the two GATA motifs within the XXIII fragment. In addition, we showed that GATA-1 could also bind specifically to a cis-acting element in the promoter of the GAPA gene, which is coordinately regulated by light with the GAPB gene. Based on these results, we propose that light controls the expression of GAPA and GAPB genes in part by regulating the binding of the same transcription factor at their GATA motifs.

Keywords: Arabidopsis; Light regulation; GATA factors; Transcription; Photosynthesis

Transcription is the primary step at which light regulates gene expression in plants [1,2]. Analyses of promoter regions of several light-inducible genes, such as RBCS, LHCb, and CHS, have identified a number of cis-acting elements that are essential for light responsiveness. In the RBCS gene family, two cis-acting elements Box II and Box III, which are essential for light responsiveness, are binding sites for the nuclear factor GT-1 [3,4]. In addition, the I- and G-boxes present in some RBCS promoters were also important for promoter activity [5]. On the other hand, the GATA and CCA motifs are the most common motifs in all known Arabidopsis LHCb1 genes [6,7]. The nuclear factor CGF-1 was shown to interact specifically with the triple GATA repeats, while a Myb-related transcription factor CCA-1 interacted with the CCA motif [6,8]. One important observation gathered from these data is that the cis-acting elements described above are not detected in many light-regulated genes. Even among members of the same gene family or between very closely related genes, cis-acting elements differ greatly from one promoter to another [1,9]. In addition, the minimum promoter sequences required to confer light responsiveness on a heterologous non-light-regulated basal promoter consist of several cis-acting elements [6,10–13]. It was suggested that combinatorial cis-acting elements are required to confer light responsiveness of light-regulated promoters in plants [1,9,11,13].

Based on sequence analysis, Grob and Stüber [14] identified a sequence motif 5′-ATGATAAAGG-3′ that is present in many LHC and RBCS genes from a number of plant species. They postulated that this element is the common light responsiveness element (LRE) shared by all light-regulated genes in which the phytochrome-mediated pathway is involved. A survey of promoter sequences of 137 light-regulated genes by Arguello-Astorga and Herrera-Estrella [9] revealed that a set of...
related modules, which contain core GATA sequence motifs, indeed existed in promoters of many, if not all, light-regulated genes. These proposed light responsive elements include the I box, GAF sites, and CGF sites in the promoters of RBCS and LHCB genes from pea and Arabidopsis [3,5,8,9]. However, the majority of the putative LREs have not been functionally characterized, and none of the transcription factors that interact specifically with these elements have been identified.

We previously showed that the nuclear genes GAPA and GAPB, which encode A and B subunits of the chloroplast glyceraldehyde-3-phosphate dehydrogenase from Arabidopsis thaliana, are coordinately regulated by light at the transcriptional level [15,16]. We identified a cis-acting element, designated as the GAP box, with the consensus sequence 5’-CATAAGA(A/G)A-3’, that is essential for light responsiveness of both GAPA and GAPB genes in transgenic tobacco plants [12,17,18]. We also showed that a nuclear factor (GAPF) from tobacco bound specifically to the GAP boxes in gel mobility shift assays [12,18]. Recently, by saturation linker-scan mutagenesis of promoter constructs in transgenic Arabidopsis, we identified additional cis-acting elements in the promoter of GAPB [19]. Mutations in one of these elements, named XXIII box, abolished light responsiveness completely, whereas mutations in other elements reduced levels of light induction. In the current study, we report that an Arabidopsis GATA-1 factor, which belongs to a class of C2C2 zinc finger transcription factors, can interact specifically with the XXIII box in vitro.

Materials and methods

Classification of Arabidopsis GATA factors. A BLAST search of the Arabidopsis genome sequence was performed using as query the amino-acid sequence of the C2C2 zinc finger of the mouse GATA-1 factor (Accession No. AAD23685). Twenty-five open reading frames (ORFs) with significant homology were identified. Alignment of the DNA binding domains of these putative GATA factors was performed using the ClustalW program. A phylogenetic tree based on the amino-acid sequences from the putative DNA binding domains of these factors was constructed using the parsimonious method in the PHYLIP version 3.57c [20].

Cloning of GATA-1 and GATA-20. The cDNA fragments encompassing the complete ORFs of GATA-1 and GATA-20 were generated by RT-PCR. To facilitate cloning, Neol and Xhol linker sequences were added to the 5’ end of the forward and reverse primers, respectively. The primer pairs used were: GATA-1, forward primer, 5’-ACCATGAGAAT GGAATCTATTCAUG-3’, reverse primer, 5’-CTCGAGTTATCCA CAACCTTTCCGATCA-3’; GATA-20, forward primer, 5’-ACCAGT GGTACCCAAAACATCTTACA-3’, reverse primer, 5’-CTCGAGTGA CACTCCTTCCGACATA-3’. After PCR amplification, the cDNA fragments were digested with Neol and Xhol and cloned into the Neol-Xhol sites of the pET32a vector (Novagen) to create in-frame amino-terminal fusions with a 6-His tag. The identities of the resulting clones, GATA-1/pET32a and GATA-20/pET32a, were confirmed by DNA sequencing.

Purification of His-tagged proteins. The GATA-1/pET32a, GATA-20/pET32a, and pET32a were transformed into Escherichia coli strain BL21(DE3). Selection of colonies that expressed optimal levels of fusion proteins and the determination of the optimal induction time were performed according to the procedures provided by the manufacturer. The induced His-GATA-20 fusion protein and the His-tag encoded by the pET32a could be detected in the soluble fraction of E. coli extracts, whereas the His-GATA-1 fusion protein was mainly detected in the inclusion bodies. The His-GATA-20 and His-tag polypeptides were accordingly purified under native conditions by the standard Ni- affinity column chromatography.

The His-GATA-1 was purified under denaturation conditions and then renatured according to the procedure described by Thanos and Maniatis [21]. Briefly, the bacteria were grown at 37°C in 50 ml LB medium containing 100 μg/ml ampicillin in a shaking incubator until the OD600 reached 0.4. IPTG was added to a final concentration of 1 mM and the incubation was continued for 3 h. The cells were harvested by centrifugation in a Beckman JA20 rotor for 20 min at 4°C. The cell pellets were resuspended in 20 ml Buffer A (8 M urea, 100 mM sodium phosphate, pH 8, 10 mM Tris, 10 mM imidazol, 0.5 mM PMSF, 10 mM 2-mercaptoethanol, and 10% glycerol) and incubated at room temperature with shaking for 20 min. The lysates were then sonicated and centrifuged at 10,000 rpm at room temperature for 20 min. Supernatants were collected and incubated with 0.5 ml of 50% Ni2+-His-Bind resin (Novagen) containing Noniident P-40 (0.5%). After 2 h, the resin was collected by centrifugation at 10,000 rpm for 2 min and washed five times with Buffer A. The bound proteins were eluted with a step gradient of 20 mM to 1 M imidazol in Buffer A. The protein fractions were analyzed by SDS-PAGE. The fractions containing His-GATA-1 protein were pooled and dialyzed against a buffer containing 6 M urea, 500 mM NaCl, 20 mM Heps, pH 7.9, 10% (v/v) glycerol, 1 mM DTT, 0.5 mM PMSF, 0.1% NP-40, and 10 μM ZnSO4. Dialysis was continued with several changes of buffers having decreasing concentration of urea as described in Thanos and Maniatis [21].

Gel mobility shift assays. The 39 bp XXIII DNA fragment was generated by the annealing of two complementary oligonucleotides and purified by PAGE. The resulting DNA fragment was end-labeled with γ-32P and T4 polynucleotide kinase for use as a binding probe. Binding reactions (25 μl volume) contained 20,000 cpm end-labeled probe, 4 μg poly(dI-dC)-poly(dI-dC), 35 mM KCl, 15 mM Heps, pH 7.6, 0.4 mM EDTA, 1 mM DTT, 0.5% (v/v) glycerol, and specific competitor DNAs as indicated in figure legends. Reactions were carried out at room temperature for 10 min with poly(dI-dC)-poly(dI-dC) and/or other competitors in the absence of labeled probe and then incubated for an additional 20 min after adding labeled probes. Re- action products were electrophoresed on 4% non-denaturing poly- acrylamide gels (acylamide:bisacylamide ratio 40:1) in 0.3× TBE buffer (1× TBE is 89 mM Tris–HCl, pH 8.0, 89 mM boric acid, and 2 mM EDTA, pH 8.0). Gels were run at 18 mA with buffer recirculation. Gels were dried onto Whatman 3 MM paper or equivalent and exposed to Kodak XAR-5 X-ray film with DuPont Lightning Plus intensifying screens at −70°C.

Results

The six cis-acting elements that are involved in light regulation of the GABP gene are illustrated in Fig. 1. In earlier studies, we showed that mutations in the XXIII box, which contains two TATC (reversed GATA) motifs, abolished light induction of GUS reporter in transgenic Arabidopsis plants [19]. Since GATA motifs are binding sites for a class of C2C2 zinc finger transcription factors in animals and fungi [22–25], we decided to investigate whether an Arabidopsis factor of similar nature could bind specifically to these motifs.
A search of the Arabidopsis genome identified 25 putative GATA genes (Table 1) with coding sequences containing the signature Cx_3Cx_17Cx_5 zinc finger. A comparison of peptide sequences of these factors revealed similarities in their putative DNA binding domains, which consist of the zinc fingers and their tail regions (Fig. 2). In contrast, outside the DNA binding domains there is no sequence homology among most of these factors, suggesting that each of these factors has a distinct activation domain. We subsequently performed a phylogenetic analysis using the amino-acid sequences of the putative DNA binding domains of these factors. The resulting parsimonious tree grouped these factors into three major classes (Fig. 2), to determine which class of these factors can bind efficiently to the GATA motifs in the XXIII box.

**Expression of His-tagged GATA-1 and GATA-20 factors in E. coli**

DNA fragments encompassing the ORFs of GATA-1 and GATA-20 were generated by RT-PCR and cloned into the pET32 vector as amino-terminal fusion with a 6-histidine tag. These vectors were transformed into the E. coli BL21(DE3, pLysS) strain and checked for IPTG-induced expression of His-tagged GATA-1 and GATA-20 proteins. We found that the induced His-GATA-20 fusion protein could be detected in the soluble fraction of E. coli extracts (Fig. 3, lane 9), while the His-GATA-1 fusion protein was mainly detected in the inclusion bodies (lane 6) and very little in the soluble fraction (lane 5). Fig. 3 also showed that the His-tagged protein produced by the pET32a vector could be detected in the soluble fraction (lane 2).

We isolated His-GATA-20 fusion proteins and the His-tag peptides encoded by pET32a by standard Ni-affinity column chromatography. The eluted His-GATA-20 was about 50% pure and the His-tagged peptide was greater than 90% pure (Fig. 3, lanes 10 and 3). His-GATA-1 protein was purified under denaturing conditions by a modified Ni-affinity column chromatography and renatured by a series of dialysis against decreasing concentrations of the denaturant [21]. The renatured His-GATA-1 was greater than 80% pure (lane 7).

**GATA-1 factor binds specifically to XXIII box in vitro**

We next performed gel mobility shift assays using a 39-bp XXIII fragment (Fig. 1) labeled with 32P as binding template and purified His-tagged GATA-1 or GATA-20 factor as sources of binding proteins. No specific binding could be detected when the GATA-20 factor or His-tag-protein encoded by pET32a was used as the source of binding protein (Fig. 4, lanes 1–3 and 4–6). In contrast, when the GATA-1 factor was used, two binding complexes could be detected (Fig. 4, lanes 8–11).
The amounts of both binding complexes increased with increasing amounts of GATA-1 factor. The formation of both binding complexes was abolished when a 200-fold excess of unlabeled XXIII was included as a competitor in the binding reaction (lane 12). These results indicated that GATA-1 factor, but not GATA-20 factor, can bind specifically to the XXIII fragment in vitro.

To determine the sequence requirement for the binding of GATA-1 factor, two modified XXIII fragments XXIII-m1 and XXIII-m2, in which the first two nucleotides of each of the two GATA motifs were replaced with CC, respectively, were generated (Fig. 1). These DNA fragments were used as binding probes in gel mobility shift assays (Fig. 5). In reactions containing 0.5, 1, or 2 pmol GATA-1 proteins with XXIII-m1 (lanes 2–4) or XXIII-m2 (lanes 8–10) as binding probes, there was no detectable binding activity. In the presence of 4 pmol GATA-1 proteins, weak binding activities could be detected in reactions with either XXIII-m1 (lane 5) or XXIII-m2 (lane 11) as binding probes. Both binding activities were abolished when 25× excess of
unlabeled XXIII fragments was used as a competitor. These results indicated that GATA-1 factor binds specifically to the two GATA motifs within the XXIII fragment and that both GATA motifs are required for efficient binding.

**GATA-1 factor binds specifically to the Al box of GAPA**

Our prior results indicated that temporal and spatial expression patterns of GAPA and GAPB genes in response to different light spectra and fluences are almost identical [15,16], suggesting that these two genes might share common cis-acting light responsive elements. Examination of the promoter sequence between Gap boxes and TATA box of the GAPA gene revealed a region located between -47 and -66 that contains a TATC (reversed GATA) motif and an AATC (reversed GATT) motif (Fig. 6A), the latter having been shown to be a binding motif for some GATA factors in animals [26,27]. The 8-nucleotide spacing between these two motifs is the same as the spacing between the two GATA motifs of the XXIII box. We decided to examine whether this GAPA promoter region, designated as the Al box, contains binding motifs for GATA-1 factor. Gel mobility shift assays were performed using the Al DNA fragment as a binding probe. As shown in Fig. 6B, little binding activity could be detected in reactions with low amounts of GATA-1 factor (lanes 2–4). However, when 4 pmol GATA-1 factor was present in the binding reaction, specific DNA–protein complexes could be detected (lane 5). These binding activities were abolished when 100× excess of unlabeled Al fragment was included as a competitor. We next synthesized two modified Al fragments, Al-m1 and Al-m2, in which the first two nucleotides of each of the two GATA motifs were replaced with CC, respectively. Gel mobility shift assays showed that mutations in m1 and m2 abolished the binding of GATA-1 factor to the Al fragment (Fig. 6C). These data indicated that both GATA and GATT motifs within the Al fragment are required for the binding of GATA-1 factor.
Discussion

GATA factors are transcriptional regulatory proteins that play critical roles in the differentiation of multiple cell types in both vertebrates and invertebrates [23,24,26,27]. A small gene family that encodes GATA factors comprised of two zinc fingers with a structure of Cx1,Cx17,C in the carboxyl domain. These factors were divided into three classes based on the amino-acid identity of the DNA binding domain (Fig. 2). Our results from gel mobility shift assays indicated that the GATA-1 factor, which is a member of the Class I factors, binds specifically to the XXIII element of GAPA (Fig. 4). The results showed that at lower amounts of GATA factor, a fast migrating DNA protein complex (FB) could be detected (lanes 8 and 9, Fig. 4). In reactions with increased amounts of GATA-1 factor, the amounts of FB increased proportionally and an additional complex with slower mobility (SB) could also be detected. Two explanations can be offered for these observations: First, GATA-1 factor can exist either as a monomer or as a dimer. The FB represents the complex formed between a GATA-1 monomer and a GATA motif, whereas the SB represents the complex formed between a GATA-1 dimer and a GATA motif. Second, GATA-1 factor exists only as a monomer. The SB band represents binding of two GATA-1 monomers to each GATA motif in the XXIII fragment. We are in favor of the second interpretation. If the first interpretation were correct, we would expect the existence of an additional binding complex with binding of two GATA-1 dimers to both motifs. In addition, it has been shown that GATA factors function as monomers in mammalian systems, despite the fact that in some cases in which tandem repeats of GATA motifs are required for efficient binding of GATA factors in vitro [32]. Nevertheless, our results showed that both GATA motifs in the XXIII fragment are required for efficient binding of GATA-1 factor, since mutations in either motif reduced the formation of both FB and SB drastically (Fig. 5).

We also showed that GATA-1 binds specifically to the GATA and GATT motifs in the Al box of GAPA (Fig. 6), although with a lower affinity than its binding to XXIII box. This could be due to the difference in nucleotide sequences outside the GATA motifs, which are known to be essential for efficient binding of GATA factors in vertebrates and fungal systems [24,26,27,32]. Alternatively, it could be due to the requirement of cofactors, since it was shown that vertebrate GATA-factors often require different classes of co-activators (Friends of GATA; FOG) to function [23,24,33].

The results we have presented indicated that the GATA motifs within the XXIII region of the GAPB and the Al region of the GAPA are binding sites for the GATA-1 factor. These observations support the idea that the promoter regions of GAPB and GAPA contain cis-acting elements that are targets of the same transcription factor. However, since there are 14 Class I GATA factors with highly conserved DNA binding domains encoded in the Arabidopsis genome (Fig. 2), it remains to be determined which GATA factor functions in vivo to mediate light regulation of GAPA and GAPB genes.

Many photosynthetic genes, including the RBCS and LHCb gene families, exhibit similar temporal and spatial expression patterns in response to different light spectra and fluences [1]. A number of cis-acting elements and trans-acting factors involved in light regulation of some of these genes have been identified. However, in spite of intensive studies on the light signaling mechanisms, little progress has been made in understanding how light controls the coordinate expression of a group of genes. How could the coordinate regulation of a group of genes by light be achieved? The easiest explanation would be that the same set of cis-acting elements and their cognate transcription factors are responsible for light activation of these genes. The existence of universal light responsive elements among a group of photosynthetic genes had been postulated [9,14]. However, these hypotheses were mainly based on comparisons of promoter sequences from plant species across great evolutionary spectra. A great majority of these elements have never been functionally characterized.

Our observation that the GATA-motif containing XXXIII Box is absolutely required for light responsiveness of GAPB may offer an opportunity to test the hypothesis of the existence of a common LRE among carbon fixation genes in Arabidopsis. We have found that the patterns of mRNA accumulation during light induction for several carbon fixation genes, including GAPA, nuclear genes encoding ribulose-1,5-bisphos-
phate carboxylase (RBCS1A), triosephosphate isomerase (TIP), and phosphoglycerate kinase (PGK), were identical to those of GAPB [15,16, and unpublished data]. It is therefore likely that the expression of these genes is regulated by light through a similar mechanism. A search of proximal promoter sequences of these genes revealed the existence of cis-elements similar to GAP boxes, AE, PI, PI, and T boxes of GAPB (Fig. 1) in some, but not all, of these genes. However, all these promoters contain GATA motifs, either as a tandem repeat (as in the case of XXIII) or as a monomer with another known cis-element located nearby (data not shown). These structures are hallmarks of GATA-factor binding sites in mammalian systems [23,24,32]. This raised the possibility that these promoter elements are common LREs for this group of carbon fixation genes. Our current results indicated that that two of these elements, XXIII box of GAPB and AI element of GAPA, contain binding sites for the same transcription factor and that light coordinates the expression of these genes in part by regulating the capability of this transcription factor to bind to the GATA motifs or to interact with other factors within the transcription complex.

References


