Expression and localization of GhH6L, a putative classical arabinogalactan protein in cotton (*Gossypium hirsutum*)

Yanfeng Wu*, Wenliang Xu*, Gengqing Huang, Siying Gong, Juan Li, Yongfang Qin, and Xuebao Li*

Hubei Key Laboratory of Genetic Regulation and Integrative Biology, College of Life Sciences, HuaZhong Normal University, Wuhan 430079, China

*These authors contributed equally to this work.

*Correspondence address. Tel/Fax: þ86-27-67862443; E-mail: xbli@mail.ccnu.edu.cn

Arabinogalactan proteins (AGPs) are a large family of highly glycosylated of hydroxyproline-rich glycoproteins that play important roles in plant growth, development, and signal transduction. A cDNA encoding a putative classical AGP named GhH6L was isolated from cotton fiber cDNA libraries, and the deduced protein contains 17 copies of repetitive motif of X–Y–proline–proline–proline (where X is serine or alanine and Y is threonine or serine). Northern blotting analysis and quantitative RT–PCR results showed that it was preferentially expressed in 10 days post-anthesis (dpa) fibers and was also developmentally regulated. A promoter fragment was isolated from cotton (*Gossypium hirsutum*) by genome walking PCR. Expression of β-glucuronidase (GUS) gene under the *GhH6L* promoter was examined in the transgenic *Arabidopsis* plants; only petiole and pedicel were stained, no staining was detected in other tissues. Subcellular localization indicated that GhH6L was localized to the plasma membrane and in the cytoplasm. These data further our understanding of GhH6L as well as shed light on functional insight to GhH6L in cotton.

**Keywords**  cotton; classical arabinogalactan protein; gene expression; subcellular localization

Received: January 1, 2009      Accepted: March 5, 2009

**Introduction**

Arabinogalactan proteins (AGPs) are a large family of highly glycosylated of hydroxyproline (Hyp)-rich glycoproteins which are thought to play important roles in plant growth and development [1]. Basically, AGPs are divided into two groups depending on their protein backbone: ‘classical’ and ‘non-classical’. The classical AGPs are defined by the core protein and are rich in Hyp, Ala, Ser, Thr, and Gly. The protein backbones have three distinct domains: an N-terminal signal peptide, a central AGP domain, and a C-terminal hydrophobic domain. This C-terminal domain is predicted to be a transmembrane helix, which suggests that this kind of AGPs may be associated with the plasma membrane [2–6]. Moreover, classical AGPs can be further divided into three subgroups: lysine-rich AGPs that have a small Lys-rich region within the AGP domain, arabinogalactan (AG) peptides, and fasciclin-like AGPs (FLAs) [5, 7].

AGPs are widely distributed in the plant kingdom from lower to higher plants. They are found in different organs and many tissues, though individual AG family members displayed changing degrees of organ-specific and tissue-specific developmental expression [4]. For instance, LeAGP1 mRNA is abundant in stylar transmitting tissues [8, 9]. AtAGP30 expression can only be detected in roots [10]. AtAGP17 was strongly expressed in leaves and stems [11]. AtAGP18 was highly transcribed in roots, flowers, and stems and weakly expressed in seedlings and rosettes [12]. AGPs are localized in cell walls, plasma membranes, intercellular spaces, and certain cytoplasmic vesicles [4, 13–15]. They are thought to be involved in different aspects associated with plant growth and development, such as fertilization, somatic embryogenesis, xylem differentiation, cell division, expansion and death, cell adhesion, and signaling cascade [8, 16–26].

Cotton, which produces the most prevalent natural fibers used in the textile industry, is one of the mainstays of the global economy. Cotton fibers, or seed hairs, which are derived from the epidermis of the ovule, are
single-cell trichomes that undergo rapid and synchronous elongation during seed development. Fiber development consists of four overlapping stages: initiation, primary cell wall (PCW) formation, secondary cell wall (SCW) thickening, and maturation [27]. Fiber elongation occurs over a period of ~21 days, with exaggerated growth rates reaching peak levels at ~10–12 days post-anthesis (dpa). At ~15 dpa, cotton fibers enter a ‘transition’ phase that signals the developmental switch from PCW to SCW synthesis [28]. Furthermore, cotton fiber offers an ideal model for studying plant cell elongation and cell wall biogenesis [29].

Recently, we identified 19 cotton genes encoding FLAs. Three of them are predominantly expressed in 10 dap fiber and are also developmentally regulated [30]. Liu et al. isolated four cDNAs encoding FLAs from cotton fibers by 5’- and 3’-rapid amplification of cDNA ends. Both northern blotting and quantitative RT–PCR analysis demonstrated that they were specifically expressed in fibers [31]. These studies demonstrated that AGPs could play important roles in fiber. Here, one cDNA coding for a classical AGP designated as GhH6L was isolated from the cotton fiber libraries, since it shared high sequence homology at amino acid level and nucleotide level with GhH6, which was previously described by John and Keller [32]. Northern blotting and quantitative RT–PCR showed that it was preferentially expressed in 10 dpa fibers and also developmentally regulated. Subcellular localization indicated that it was localized in the plasma membrane and in the cytoplasm. These data lead to a better understanding of GhH6L as well as provide functional insight to GhH6L in cotton.

Materials and Methods

Plant materials and growth conditions
Cotton (G. hirsutum cvs. Xuzhou142 and Cocker 312) seeds were surface-sterilized with 75% (v/v) ethanol for 1 min and 10% (v/v) H2O2 for 1 h, followed by washing with sterile distilled water. The sterilized seeds were germinated on 1/2 MS medium under 16/8 h light/dark cycles at 28°C. The seedlings were transplanted into soil for further growth to maturation. Tissues for DNA and RNA extraction were derived from cotton plants grown in a greenhouse.

Isolation of GhH6L cDNA
More than 4000 cDNA clones were randomly selected from cotton fiber cDNA library [33] for sequencing. One clone containing the complete GhH6L sequence was identified from the cDNAs.

DNA and protein sequence analysis
Unless otherwise stated, nucleotide and amino acid sequences were analyzed using DNASTar. Promoter prediction software (http://www.fruitfly.org/seq_tools/promoter.html) was used to predict the GhH6L transcription initiation site. Putative cis-acting elements were identified by submitting the promoter sequence to the PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Identification of protein domains and significant sites were found by using Motifscan (http://myhits.isb-sib.ch/cgi-bin/motif_scan); The hydropathic profile was calculated according to Kyte and Doolittle hydropathy plot (http://fasta.bioch.virginia.edu/fasta_www/grease.htm); SignalP (www.cbs.dtu.dk/services/SignalP/) was used to determine the N-terminal signal sequence, potential N-glycosylation sites were found by using NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/). The glycosylphosphatidylinositol (GPI) Plant Prediction Server (http://mendel.imp.ac.at/gpi/plant_server.html) was used to predict GPI anchor site.

RNA extraction, purification, and reverse transcription
Total RNA was extracted from roots, hypocotyls, cotyledons, leaves, petals, anthers, ovules, and developing fibers (2–20 dpa) as described previously [33]. RNAs were purified using Qiagen RNAeasy mini kit (Qiagen, Hilden, Germany) according to the manufacturer’s instruction. First-strand synthesis of cDNA was performed using M-MLV reverse transcriptase (Promega, Madison, USA) according to the manufacturer’s instruction.

Northern hybridization
The 3’-untranslated region (3’-UTR) fragment of GhH6L cDNA was prepared as a probe by PCR amplification. The sense primer is 5’-CAGCTGGAACTGA CACGAGT-3’, whereas the reverse primer is 5’-GTAACGTAGGCTCTCTAGTAG-3’. RNA samples (20 μg per lane) from the cotton tissues were separated on 1% (w/v) agarose-formaldehyde gels for 4–5 h and then transferred onto Hybond-N+ nylon membranes (Amersham, Buckinghamshire, UK) by capillary blotting. Gene-specific probe was labeled with [a-32P]dCTP using the random primer method (Random primer DNA labeling kit; TaKaRa, Dalian, China). Northern hybridization was carried out as described previously [33].
The membrane was exposed to X-film (Eastman Kodak, Rochester, USA) with two intensifying screens at −70°C for 1–3 days.

**Quantitative RT–PCR**

Expression of GhH6L as well as GhH6 in cotton tissues was analyzed by quantitative RT–PCR using the fluorescent intercalating dye SYBR-Green (TOYOBO, Tokyo, Japan) in a detection system (Opticon2; MJ Research). A cotton polyubiquitin gene (GhUBI) was used as a standard control in the RT–PCR reactions. A two-step RT–PCR procedure was performed in all experiments using the method described previously [30]. The primers used in RT–PCR are:

GhH6L-up 5′-CAGCTGGAAAACAGACACGAG-3′,
GhH6L-down 5′-GTAACGTAGGCTCTAGTAG-3′;
GhH6-up 5′-CAGCTGGAAAGACACGAGC-3′,
GhH6-down 5′-TGTAGGCTCTCATATCGAT-3′.

**Isolation of GhH6L promoter by genome walking and GUS reporter vector construction**

GhH6L 5′-flanking sequence was isolated according to the Genome walker kit (Clontech, Mountain View, USA) as described previously [33]. A 1688-bp promoter fragment of GhH6L was amplified by PCR, and a SalI site and a BamHI site were introduced at the 5′-terminal and 3′-terminal of the GhH6L promoter, respectively. The SalI/BamHI fragment of GhH6L in 5′-upstream region was subcloned into the SalI/BamHI sites of the pBI1101 vector to generate the chimeric GhH6L:GUS construct.

**Arabidopsis transformation and GUS assay**

For transformation, 200 ml of YM liquid medium containing 50 μg/ml of kanamycin and 100 μg/ml of streptomycin was inoculated with 2 ml overnight culture of Agrobacterium tumefaciens LBA 4404 harboring the GhH6L::GUS construct and was grown for an additional 20 h at 28°C until OD₆₀₀ reached 0.8. Cells were harvested by centrifugation, resuspended in 400 ml of 5% sucrose solution containing 0.02% (v/v) of silwet L-77 (Lehle seeds), and used for transformation, which was carried out using the floral-dip method [34]. Seeds were harvested and stored at 4°C. For screening, seeds were sterilized in 75% (v/v) ethanol for 1 min and then in 10% NaClO solution for 10 min, followed by four washes with sterile water. Kanamycin-resistant plants were selected by incubating plants on MS salt plates containing 50 μg/ml of kanamycin.

Histochemical assays for GUS activity in transgenic plants were conducted according to the protocol previously described by Jefferson et al. [35] with some modifications. In brief, fresh tissues from the plants were incubated in 5-bromo-4-chloro-3-indolylglucuronic acid (X-gluc) solution consisting of 0.1 M sodium phosphate (pH 7.0), 10 mM ethylenediaminetetraacetic acid, 0.5 mM potassium ferrocyanide, and 0.1% (w/v) X-gluc for 8–12 h. The stained plant materials were then cleared and fixed by rinsing with 70% (v/v) ethanol, and the samples were examined and photographed under a Leica stereomicroscope (Leica MZ16f, Leica, Germany).

**Subcellular localization of GhH6L**

GhH6L::eGFP (enhanced green fluorescence protein) was constructed by excising the coding region of GhH6L with BamHI and XbaI and cloned into the binary vector pBI121. Also, eGFP coding region was excised with XbaI and SacI and cloned into pBI121 downstream of GhH6L. The resulting constructs were introduced into A. tumefaciens LBA 4404 and transformed into cotton as described previously [33], transgenic cotton callus cells were observed under a Leica Confocal laser scanning fluorescence microscope (Leica TCS SP5).

**Results**

**Isolation and characterization of GhH6L**

To isolate genes involved in the development of cotton fiber, over 4000 cDNA clones from the fiber cDNA library were randomly sequenced. One cDNA encoding a putative AGP was selected and designated as GhH6L (GhH6-like; accession number in GenBank: FJ600363). This cDNA encoded 214 amino acids (20.66 kDa) and is proline-rich, containing 17 copies of repetitive motif of X–Y–proline–proline–proline (where X is serine or alanine and Y is threonine or serine) typical of cell wall proline-rich protein. When protein blast was performed in the GenBank database using GhH6L as a query, only three proteins have sequence similarity with GhH6L. The three proteins are GhH6, GbH6L, and AtAGP18. GhH6L has 99% amino acid identity with GhH6 and GbH6L (that is why we named the gene as GhH6L) and 34% identity with AtAGP18. Between GhH6L and GhH6, only two amino acid residues are different. P was replaced by T at position 38 and L was replaced by P at position 60 (Fig. 1). AtAGP18 is one of lysine-rich classical AGPs, except for the lysine-rich region, which all have the same domain organization: an N-terminal signal...
peptide, a central AGP domain, and a C-terminal GPI anchor addition sequence [Fig. 2(A)]. For GhH6L, the most likely signal peptide cleavage site is between positions 23 and 24 (VLG23 –Q24A), and best predicted GPI addition site is S at position 188. To evaluate if the gene contains any intron sequences, PCR was employed to amplify the full-length sequence of GhH6L. The PCR products obtained from both cvs. Xuzhou142 and Cocker 312 (Fig. 3) of cotton (G. hirsutum) were cloned and sequenced, demonstrating that all of two cotton cultivars contain the same copy of the GhH6L gene.

Compared with its cDNA sequence, a 580-bp intron was found in the open reading frame (ORF) of the GhH6L gene [Fig. 2(B)]. Because the 5'-UTR sequence was not available in GenBank, we then compared the coding region and 3'-UTR sequences of GhH6L and GhH6. Though there are only five interchanges in the coding region, the 3'-UTR shows more divergence and the identity is ~89% (Fig. 4), indicating that they are not the same gene.

GhH6L gene is preferentially expressed in fiber

To investigate the expression pattern of the GhH6L gene, RNA gel blotting was carried out using total RNAs isolated from the different cotton tissues. Because GhH6L shares high nucleic acid sequence similarity to GhH6, the probe selection is critical for the experiment. Since these two genes show more divergence at the 3'-UTR, one reverse primer specific to GhH6L was selected. Nucleotide blast was performed using this fragment as a query found no similar sequences in cotton (G. hirsutum), indicating it is specific to GhH6L. We are confident that northern blotting and quantitative RT–PCR results are reflecting the expression of GhH6L rather than other similar genes.

Northern blotting analysis showed that the GhH6L gene was preferentially expressed in fibers; much smaller
but not negligible amounts of its transcripts were detected in ovules, and no transcript was detected in other tissues analyzed (Fig. 5). Quantitative RT–PCR also revealed that the $GhH6L$ gene was predominantly expressed in fibers, whereas much smaller or negligible amounts of its transcripts were detected in other tissues (such as ovules, anther, etc.), which was quite consistent with the northern blotting analysis. To investigate whether the $GhH6L$ gene expression is developmentally regulated in fibers, quantitative RT–PCR was also used for further analyzing the expression pattern of the $GhH6L$ gene during fiber development. Meanwhile, the $GhH6$ transcripts were also detected during fiber development in the same cultivar of cotton (Fig. 6). We were unable to isolate RNA...
from 25 dpa or later when cellulose was being deposited. These results revealed that the GhH6L gene transcripts were accumulated at relatively high levels in all stages of fiber elongation. The transcription appeared to increase in 15 dpa fiber and had a little decrease in 20 dpa fiber. All these showed that the expression of GhH6L was fiber-preferential and developmental-regulated, suggesting that it may participate in fiber development.

Isolation and activity analysis of GhH6L promoter

On the basis of the cDNA sequence of GhH6L, a 1688-bp promoter fragment was isolated from cotton (G. hirsutum) by genome walking PCR. Analysis of the isolated 1688-bp upstream of the GhH6L ATG translation start codon revealed a putative transcription initiation site 91-bp upstream of the ATG translation start codon (http://www.fruitfly.org/seq_tools/promoter.html), and TATA and CAAT boxes were found in this promoter. In addition, several putative cis-acting regulatory elements were identified, such as light-response elements (ATTAATTTTACA, AATCTAATCT, ATTAAT, TTTCAAA, CACGTC, and GGTTAAT), the anaerobic induction element (TGGTTT), gibberellin-responsive element (TCTGTTG and TATCCCA), heat stress responsiveness (AAAAATTTTC), ethylene-responsive element (ATTTCAAA), and an element required for endosperm expression (GTGAT) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Since hormone and/or stress-response elements are present in the promoter region, it is possible that GhH6L expression may be important for plant development.

Because generating transgenic cotton is time-consuming and labor-intensive, to overcome the difficulties of regenerating cotton, heterologous systems were used to determine if fiber-specific promoters drove leaf trichome-specific expression in other plants. Studies of these promoter regions in plants such as Arabidopsis and tobacco have shown that cotton fibers and Arabidopsis or tobacco trichomes may share common regulatory elements for tissue-specific expression. So we took advantage of Arabidopsis for promoter activity analysis. Expression of the GUS gene under the control of GhH6L promoter was examined in transgenic Arabidopsis plants. In 7- and 14-day-old seedlings, the staining was only detected in petiole and was intense in the midvein region of leaves [Fig. 7(A,B)]. After transition to the reproductive growth phase, GUS staining was restricted in pedicels [Fig. 7(C)] besides in the petiole of rosette leaves [Fig. 7(D)], and no expression was detected in other tissues analyzed in this study.

Subcellular localization of GhH6L protein

To test the localization of GhH6L protein in cotton cells, GhH6L-eGFP fusion protein was expressed in transgenic cotton cells under the control of 35S CaMV
promoter. The eGFP fluorescence was monitored under a Leica Confocal laser scanning microscopy. Wide-type cotton callus cells showed no background fluorescence, whereas green fluorescence was observed on the surfaces (cell wall or plasma membrane) as well as in cytoplasm of the transgenic cotton cells [Fig. 8(A,B)]. However, when the cells were plasmolyzed with 4% NaCl for 15 min, green fluorescence was localized in the cytoplasm and the plasma membrane [Fig. 8(C,D)], indicating that the GhH6L protein may be localized to the plasma membrane and in the cytoplasm.

Discussion

Cotton H6 (GhH6), as addressed by John and Keller [33], is more like an AGP because the predominant amino acids are Pro, Ala, Ser, and Thr. Furthermore, GhH6 is readily soluble in low-salt buffers. In this paper, a cDNA predominantly expressed in fiber was isolated and characterized, and the deduced amino acid sequence shared 99% similarity with GhH6 [32], which implied that these two genes might originate from a series of gene duplication during evolution and they most likely belonged to the A genome in the allopolyploid cotton (G. hirsutum with A and D genomes). In addition, our results revealed that GhH6L might localize to the plasma membrane and cytoplasm. Several lysine-rich AGPs were found to exist on the plasma membrane and hechtian strands, such as LeAGP1, AtAGP17, and AtAGP18 [11, 22, 25]. This localization may be determined by a GPI anchor. Many GPI-anchored glycoproteins have been identified [6]. All of the four FLAs isolated by Liu et al. [31] have GPI anchor sites, whereas 15 of the 19 GhFLA proteins isolated in our lab have GPI anchor sites [30]. Likewise, GhH6L also has a GPI anchor site at its C-terminus. It seems that GPI anchors may be a common feature in classical AGPs. Udenfriend and Kodukula [36] experimentally determined the GPI attachment site (designated ω). Only Ser, Asn, Ala, Gly, Asp, and Cys are found at ω site, and only Ala, Gly, Thr, or Ser is found at the ω + 2 site. GhH6L (ω = Ser188 and ω + 2 = Ala190) applies to this pattern. According to the above features that GhH6L has, we propose that GhH6L encodes a putative classical AGP.

Though GhH6L and GhH6 have extraordinary high amino acid sequence similarity, expression patterns of GhH6L and GhH6 were different. For example, GhH6L highly expressed in fiber cells through 2–20 dpa fibers, whereas GhH6 mRNA is specifically restricted in fiber cells through 10–24 dpa fibers [32]. Furthermore, GhH6L expression was found in 10 dpa ovules, though the transcription level was very low; real-time RT–PCR experiments revealed that even lower levels of transcripts were found in anthers and hypocotyls. The expression patterns of AGPs provided hints in elucidating their biological roles. Since GhH6L is very active during rapid fiber cell elongation stage even as developmental stages switch from PCW to SCW, we suppose it may play important roles especially during early fiber developmental stages.

With the isolation and characterization of a fiber-specific promoter, we are able to address the molecular mechanisms underlining the fiber development and express target gene products in the developing fiber through genetic engineering. Promoters of several fiber specific genes have been investigated [33, 37–39]. However, cotton transformation mediated by Agrobacterium is time-consuming and labor-intensive to obtain transgenic plants. Because trichomes in Arabidopsis and tobacco show a lot of structural and genetic similarities with cotton fibers, many researchers use these heterologous model species to monitor the activity of cotton fiber-specific promoters. Very recently, a promoter of a cotton fiber-specific MYB gene directs expression specific to Arabidopsis trichomes and tobacco glandular trichomes [40]. Promoter of another fiber-specific gene encoding lipid transfer protein named Fsltp4 was strongly active in tobacco leaf trichomes [41]. These results suggest that some general parallels may regulate trichome morphogenesis in Arabidopsis, tobacco, and cotton. On the other hand, GUS activity analysis revealed that the promoter of GhGlcAT1 which

Fig. 8 GhH6L was localized to the plasma membrane and cytoplasm (A) Epifluorescence of the cells. (B) The clear field image of the same cells shown in (A). (C) Epifluorescence of cells plasmolyzed with 4% NaCl for 10 min. (D) The clear field image of the same cells shown in (C). Scale bars = 50 μm.
predominantly expressed in 15 dpa fiber drives efficient expression of the GUS gene in the root cap, seed coat, pollen grains, and trichomes [39]. Promoter of GhRGP1, which preferentially expressed in fiber cells, was fused to the GUS gene and GUS activity was analyzed. The highest expression level under the control of the GhRGP1 promoter was observed in 4-week-old transgenic tobacco roots, lower GUS activity was detected in stems and leaves. In the reproductive growth stage, the GUS expression was detected in pistils, stigmas, styles, ovary, anthers and trichomes, seeds, and vascular tissues of stems [38]. Our data demonstrated the promoter of GhH6L directs expression in petioles and pedicles of transgenic Arabidopsis plants. These results strongly suggest that fiber and trichome specificity are regulated by direct mechanisms in cotton and other plants such as tobacco and Arabidopsis.

Acknowledgements

The authors are especially grateful to Brian Keppler at Ohio University for critical reading of the manuscript.

Funding

This work was supported by the grants from the National Natural Sciences Foundation of China (30870142 and 30400022) and Chenguang Project for Young Scholars in Wuhan City (20065004116-26).

References