Original

Cloning and Sequencing of Cotton Homologs of *bcsA* Gene Encoding Cellulose 4-β-Glucosyltransferase*[1]

Yuri IHARA*[2], Fukumi SAKAI*[2]
and Takahisa HAYASHI*[2]

(Received May 31, 1997)

Abstract——A cDNA clone derived from developing cotton fiber cells has been isolated as a homolog of *Acetobacter xylinum bcsA* gene encoding cellulose 4-β-glucosyltransferase. The clone characterized as *pcsA2 (celA2)* appears to be a full length cDNA of 3,311 bp and has an open reading frame of 3,120 bp that encodes a polypeptide of 1,039 amino acids with a calculated molecular mass of 125 kDa. The bacterial cellulose 4-β-glucosyltransferase shows sequence similarity to the cotton cDNA gene product at four regions, one of which is the binding site for UDP-glucose.

Keywords: pcsA2, bcsA, cellulose 4-β-glucosyltransferase, cotton fiber

1. Introduction

A high rate of cellulose synthesis starts at the onset of secondary wall formation in cotton (*Gossypium hirsutum*) fiber cells[1]. The transition of wall formation from primary to secondary occurs between 18 and 21 days post-anthesis. This transition may involve the changes in level of expression of genes coding for polypeptides required for cellulose synthesis. An cDNA library was constructed in the plasmid vector pBluescript with mRNA obtained from cotton fiber cells harvested at 18 to 21 days post-anthesis. Random sequencing of 1,000 clones from the cDNA library revealed about 200 clones with sequences found in the Gene Bank database. The identified clones contained several cDNAs for cell wall-related enzymes, endo-1, 4-β-glucanase, endoxyloglucan transferase, endo-1, 3-β-glucanase and sucrose synthase[2]. This paper describes the identification and definition of *pcsA2* for the candidate of cellulose synthase gene in cotton fiber cells.

2. Materials and Methods

2.1 Sequencing of cDNA clones

Cotton plants (*Gossypium hirsutum* L.) were grown in a growth chamber. The fibers

*[1] This work was presented at the 47th Annual Meeting of the Japan Wood Research Society at Kochi, April 3, 1997.

*[2] Laboratory of Gene Expression.

were harvested and immediately put into liquid nitrogen at different developmental stages. Total RNA was isolated from cotton fibers according to the method of Hall *et al.*[3] and poly $(A)^+$ RNA was isolated from total RNA with an oligotex-dT30 (TOYOBO). A cDNA library was constructed in the plasmid vector Bluescript by the Lambda Uni-Zap vector (Strategene) using cDNA with poly$(A)^+$ RNA of cotton fibers harvested at 18 to 24 day as the template. One thousand clones were randomly selected from the cDNA library, and the plasmid DNA inserts were sequenced using an Applied Biosystems model 373A DNA sequencer[2].

2.2 Full length cDNA preparation

A full length cDNA was obtained from the reverse transcription of total RNA with the gene-specific primer of the cDNA clone according to the 5'RACE method[4]. One mg of RNA from cotton fibers harvested at 15–21 days post-anthesis was added to 20 ml of 50 mM Tris/HCl, pH 8.5, 6 mM $MgCl_2$, 40 mM KCl, 1 mM DTT, each dNTP at 1.5 mM, 10 units of Superscript II RT (Gibco BRL) and gene-specific primer-1 (GGAATGATGAATTTGCCGG), and the mixture was incubated at 42°C for 50 min. Then, RNA was digested with 10 units of RNaseH at 37°C for 30 min, and cDNA was isolated with a Quick-spin column (Boehringer Mannheim). Poly (dC)-tail was added to the 5′ end of the cDNA using 15 units of terminal deoxynucleotidyl transferase in a total volume of 25 ml containig 0.2 mM dCTP. The tailed cDNA was amplified by PCR in a total volume of 50 ml containing the gene-specific primer-2 (AGATTGCCGGTAGGGTACAATAGGCG), abridged anchor primer (Gibco BRL), and Advantage Klentaq Polymerase mix (Clontech). PCR products were size-fractionated by electrophoresis in 1.2% agarose gel, and the desired cDNA which appeared to be 2,000 bp were excised from gel with GENO-BIND (Clontech). The cDNA was reamplified by PCR, and the products were ligated into pGEM-T vector (Promega), then the inserted region was sequenced. Full-length cDNA was obtained by PCR in the reaction mixture containing the specific primers as 5′ end and 3′ end of cDNA. The amplified DNA was cloned into pGEM-T vector.

3. Results and Discussion

Seven hundreds and fifty cDNA clones were computer-simulatively translated according to their nucleotide sequences and their potential 4,500 polypeptides in deduced amino acid sequences were subjected to homology search with *bcsA* gene product (*Acetobacter xylinum* cellulose 4-β-glucosyltransferase)[5]. Fifteen clones which exhibit homology with the amino acid sequence of the *bcsA* gene product are classified into three types, *pcsA2*, *pcsA3* and *pcsA4* in the amount of 7, 7 and 1 clones, respectively, according to the definition shown by Pear *et al.*[6] These clones show homology partially (60 to 70%) at nucleotide level. There

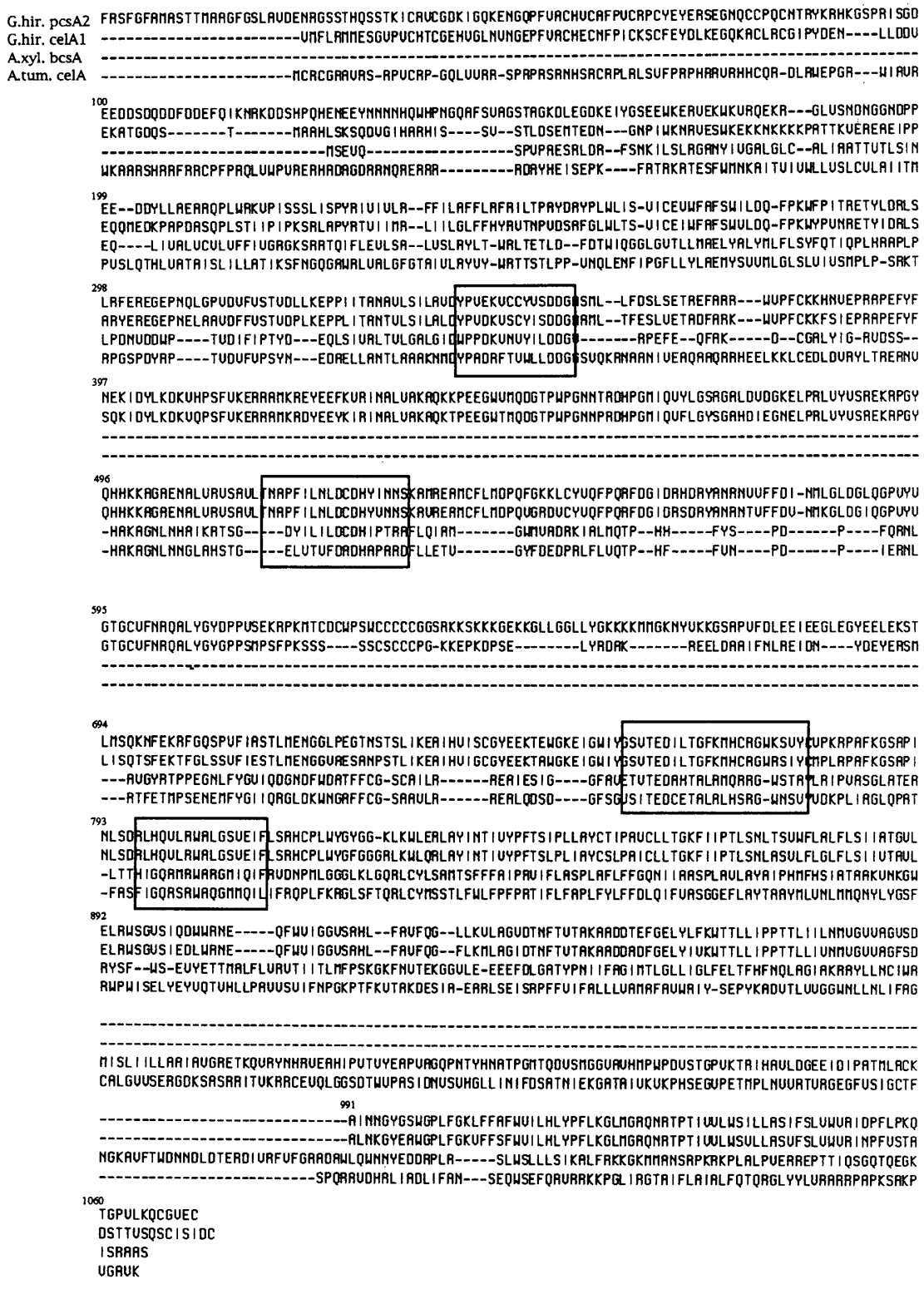

Fig. 1. Multiple alignment of deduced amino acid sequences of cotton and bacterial polypeptides. Analyses were performed by Clustal w (1.60) and DNASYS-Mac v3.0. Residues are boxed when they show similarity between the cotton cDNA and bacterial cellulose 4-β-glucosyltransferase, including the binding site for UDP-glucose.

Table 1. Characterization of *pcsA2* cDNA and its deduced amino acid sequence

Gene	Length (bp)	Identity with *bcsA* (%)		UDPG-binding motif	Source
		Nucleotide	Amino acid		
pcsA2	3,311	42.8	26.6	YPVEKVCCYVSDDG	Cotton
celA1	3,200	42.4	25.4	YPVDKVSCYISDDG	Cotton[6]
bcsA	2,262	100	100	WPPDKVNVYILDDG	*Acetobacter*[5]
celA	2,583	48.8	28.9	YPADRFTVWLLDDG	*Agrobacterium*[7]

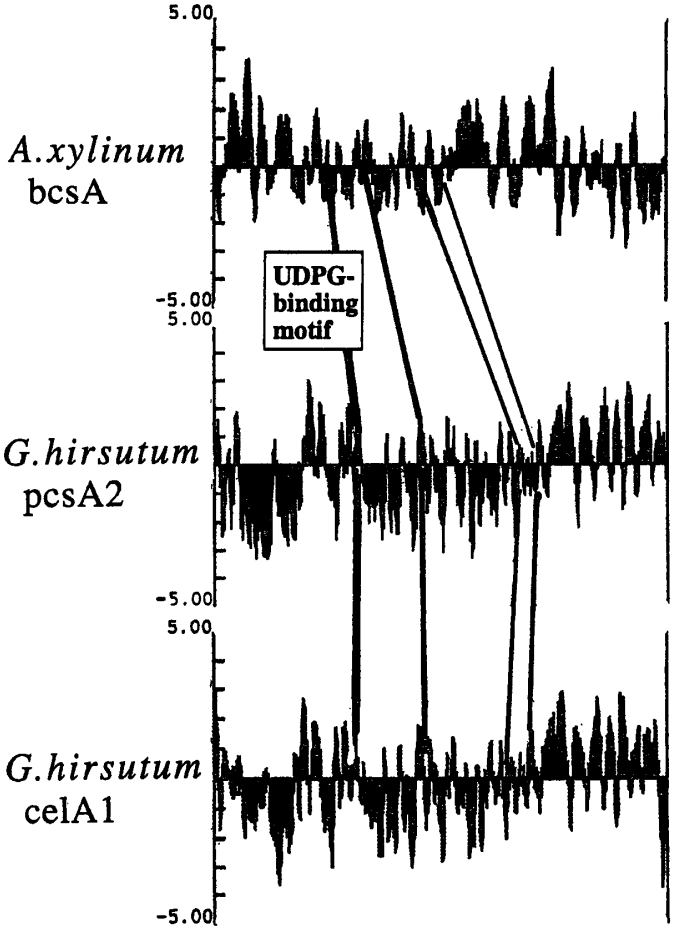

Fig. 2. Hydropathy plots of cotton pcsA2 aligned with those of bacterial bcsA and cotton celA1 polypeptides. The hydropathy profiles were calculated using a Kyte & Doolittle program and window size of 10. Four lines indicated the regions that show similarity between the cotton cDNA and bacterial cellulose 4-β-glucosyltransferase, including the binding site for UDP-glucose.

were no *pcsA1 (celA1)* in the 15 clones, although the full length cDNA of *celA1* was obtained by Pear *et al.*[6]

The full length cDNA of *pcsA2* has been obtained by generating cDNA with the PCR

technique to amplify copies of the region between a specific point in the transcript and 5′ end because *pcsA2* was one of the major clones in the homologs of our cDNA library. Fig. 1 shows the results of a multiple alignment of the deduced amino acid sequences from cotton *pcsA2*, cotton *celA1*[6], *Acetobacter bcsA*[5] and *Agrobacterium ce al.*[7]. The cotton *pcsA2* which appears to be a full length clone of 3,311 bp contains an open reading frame of 3,120 bp that encodes a polypeptide of 1,039 amino acids with calculated molecular mass of about 125 kDa. The deduced amino acid sequence contains one consensus sequence for UDP-glucose binding motif (Table 1) and two consensus sequences for a potential site of N-linked glycosylation at Asn-509 and Asn-771.

Cellulose 4-β-glucosyltransferase in *Acetobacter xylinum* exhibits 42.8% identity at the DNA level and 26.2% identity at the whole deduced amino acid level to pcsA2 polypeptide. The cotton celA1 polypeptide exhibits 53.9% identity at the DNA level and 68.7% identity at the amino acid level to the pcsA2 polypeptide. Hydropathy analysis showed that the distribution of hydrophilic and hydrophobic regions is similar between polypeptides of pcsA2 and celA1 but very different in both the N-terminal and C-terminal regions between the polypeptides of cotton and *acetobacter* (Fig. 2). Nevertheless, the bacterial cellulose 4-β-glucosyltransferase shows sequence similarity to the cotton cDNA gene product at four regions, one of which is the binding site for UDP-glucose. The hydropathy profiles also suggest at least two transmembrane helices, *e.g.*, one is located in the N-terminal region (amino acid sequence No. 215–238) and the other is in C-terminal region (amino acid sequence No. 1,007–1,031). The central regions of the polypeptides are rather hydrophilic and are probable catalytic sites in the cytoplasm. The hydrophilic regions have the conserved UDP-glucose binding motif which has been believed to bind to substrate and to catalyze the transfer of glucose into 1, 4-β-glucan.

In conclusion, we have identified the full length cDNA of *pcsA2* from an cDNA library derived from developing cotton fiber cells for the first time. Detailed analysis of Northern blots of *pcsA2* mRNA is in progress to determine in which organs and when the *pcsA2* is specifically expressed in cotton plants.

References

1) Y. SHIMIZU, S. AOTSUKA, O. HASEGAWA, T. KAWADA, T. SAKUNO, F. SAKAI and T. HAYASHI: *Plant Cell Physiol.*, **38,** 375–378 (1997).
2) O. HASEGAWA, S. AOTSUKA, F. TADOKORO, S. TAKENISHI and H. UCHIMIYA: *In* Kyoto Conference on Cellulose. Edited by N. Shiraishi, pp. 87, The Cellulose Society of Japan, Kyoto (1994).
3) T.C. HALL, Y. MA, B.U. BUCHBINDER, J.W. PYNE, S.M. SUN and F.A. BLISS: *Proc. Natl. Sci. USA,* **75,** 3196–3200 (1978).
4) M.A. FROHMAN, M.K. DUSH, and G.R. MARTIN: *Proc. Natl. Sci. USA,* **85,** 8998–9002 (1988).
5) H.C. WONG, A.L. FEAR, R.D. CALHOON, G.H. EIDHINGER, R. MAYER, D. AMIKAM, M. BENZIMAN, D.H. GELFAND, J.H. MEADE, A.W. EMERICK, R. BRUNER, B.A. BEN-BASAT and R. TAL: *Proc. Natl. Sci. USA,* **87,** 8130–8134 (1990).

6) J.R. PEAR, Y. KAWAGOE, W.E. SCHRECKENGOST, D.P. DELMER and D.M. STALKER : *Proc. Natl. Sci. USA,* **93,** 12637–12642 (1996).
7) A.G. MATTHYSSE, S. WHITE and R. LIGHTFOOT : *J. Bacteriol.,* **177,** 1069–1075 (1995).