Theor Appl Genet (1997) 94:909–918

S. J. Orford · J. N. Timmis

Abundant mRNAs specific to the developing cotton fibre

Received: 24 September 1996 / Accepted: 18 October 1996

Abstract Five fibre-specific cDNA clones were isolated by differential screening of a cDNA library from cotton fibres, a developmentally synchronous population of non-dividing cells. The genes corresponding to these cDNAs were expressed preferentially in fibre cells and exhibited differing patterns of temporal expression during fibre development. One cDNA encoded a lipid transfer protein (LTP), and a second encoded a member of a group of well-characterised proline-rich proteins (PRP) from plants. The presence of signal peptide-encoding sequences suggests that both the LTP and the PRP are targeted to the extracellular matrix of the fibre, and a role is envisaged for each in cell elongation. Sequence analysis showed that a third clone was similar to a previously reported fibre-specific sequence of unknown function, whilst the remaining two cDNA clones showed no sequence similarity to previously reported plant nucleic acids.

Key words *Gossypium hirsutum* · Differential gene expression · Lipid transfer protein · Proline-rich protein · cDNA sequence

Introduction

The commercial cotton fibre is a product of Malvaceous plants of the genus *Gossypium*. Cotton fibres arise synchronously from single cells of the outer epidermis of ovules and begin to differentiate at about the time of anthesis. The signals which trigger differentiation of an epidermal cell are unknown, but "prefibre" cells can be discerned from other epidermal cells as early as 16 h preanthesis (Ramsay and Berlin 1976). Fibre cells develop in four distinct but overlapping phases. An initiation period is followed by extensive cell elongation (primary wall synthesis) and then by secondary wall deposition, which begins 16–19 days postanthesis (DPA) and continues until the cotton fibre is mature (Basra and Malik 1984). The degree of overlap between the phases of elongation and secondary wall thickening varies between cultivars and under differing environmental conditions (Schubert et al. 1973; Jasdanwala et al. 1977). Overlap tends to be greater in long-fibred cultivars than in short-fibred cultivars (Naithani et al. 1981), and the time of onset of secondary wall deposition may affect final fibre length (Beasley 1979).

Cell-wall biosynthesis is the major synthetic activity in fibre cells. The rate of cellulose synthesis increases rapidly at 16 DPA, and cotton fibre cells are 95% cellulose at maturity (Young 1986). Specific cell-wall characteristics influence a number of textile properties in cotton. Of obvious importance are chemical composition of the wall, degree of cellulose polymerisation and amount of cellulose deposited (Rowland et al. 1976; Gupta et al. 1979; Timpa and Ramey 1989). Fibre elongation is driven by turgor pressure generated by the influx and accumulation of potassium and malate in the enlarging central vacuole (Dhindsa et al. 1975). The osmoregulation of these solutes may play an important role in the rate and extent of cotton fibre growth.

The economic importance of cultivated cotton has stimulated extensive studies of fibre and seed development. Manipulation during domestication has resulted in modern cotton varieties in which the fibre is many times longer and stronger than that of any wild ancestor. Fibre properties, particularly length, strength and fineness, require further improvement to meet economic pressures and to facilitate more efficient spinning technology (Colton 1991). Because these characters are under genetic control (Basra and Malik

Communicated by R. Hagemann

S. J. Orford · J. N. Timmis (✉)
Department of Genetics, The University of Adelaide, Adelaide 5005, Australia

1984), modification of fibre development through genetic manipulation could result in enhanced fibre quality parameters. Prior to this, an understanding of the regulatory elements and genes expressed in cotton fibres is required.

Since cotton fibres grow synchronously and are single, elongated cells which are relatively homogeneous in size, they are a suitable experimental system in which to study cellular and developmental events. In this paper we report the isolation and characterisation of cDNA clones corresponding to mRNAs which are present in fibre cells but absent in other differentiated cotton tissues.

Materials and methods

Plant material and growth conditions

Seeds of *G. hirsutum* L., cv 'Siokra 1-2' were obtained from Cotton Seed Distributors, Narrabri, NSW, sown in soil and grown in growth cabinet conditions (day/night temperatures of 30°C/25°C, light/dark cycle of 16/8 h). Fibres were manually separated from the ovule epidermis at various stages of development and frozen in liquid nitrogen. Leaf and whole flowers were collected from mature plants and stored at −70°C. Root and hypocotyl (seedling) tissue was harvested when the plants had reached a height of approximately 5 cm.

Differential screening of a cDNA library

Total RNA was isolated from all cotton tissues using the method of Wan and Wilkins (1994) and the RNA stored in 80% ethanol at −70°C. Polyadenylated [poly(A)$^+$] RNA from leaf and 13-DPA fibre was purified on oligo(dT)-cellulose columns (mRNA Purification Kit, Pharmacia) according to the manufacturer's instructions.

Double-stranded cDNA was synthesised from 13-DPA fibre poly(A)$^+$ RNA using a commercial cDNA synthesis kit, according to the manufacturer's instructions (Pharmacia). Complementary DNA was ligated into *Eco*RI-digested and dephosphorylated λZAP®II vector (Stratagene), packaged using *in vitro* packaging extracts (Promega) and plated according to Stratagene instructions. Phage were lifted in duplicate onto Hybond-N$^+$ membranes (Amersham) and differentially screened with [^{32}P]-labelled single-stranded cDNA probes from poly(A)$^+$ RNAs of leaf and 13-DPA fibre.

To prepare single-stranded cDNA probes, 2–4 μg of poly(A)$^+$ RNA in 3 μl of DEPC-treated H_2O was heated to 65°C for 5 min and transferred to ice for 5 min. After briefly microfuging the sample, the following components were added in order: 1 μl DTT (10 m*M*), 20 U RNasin® (Promega), 3.1 μl TRIS-HCl pH 8.3 (400 m*M*), 1.2 μl KCl (625 m*M*), 0.5 μl $MgCl_2$ (500 m*M*), 1.5 μl oligo(dT) (1.6 μg/μl, Amersham), 1.3 μl each of dATP, dTTP and dGTP (10 m*M*), 4.4 MBq α-[^{32}P]dCTP (Amersham) and 1 μl MLV reverse transcriptase (200 U/μl, BRL). The mix was incubated at 41°C for 2 h.

Preliminary analysis of cDNA clones

Inserts from differentially hybridising λZAP®II plaques were rescued into the phagemid vector pBluescript®SK(−) (Stratagene) for further manipulation. Plasmid DNA was purified using the alkaline lysis method of Sambrook et al. (1989). Polymerase chain reactions (PCR) across cDNA inserts were performed in 25-μl reaction mixtures containing 200 ng plasmid DNA, 1 × PCR reaction buffer (Bresatec), 0.4 m*M* each of dATP, dCTP, dGTP and dTTP, 4 m*M* $MgCl_2$, 0.7 μ*M* each of T3 and T7 primers and 1.5 U of *Taq* DNA Polymerase (Bresatec). Initial template denaturation was at 94°C for 2 min followed by 25 PCR cycles of denaturation at 94°C for 1 min, primer annealing at 45°C for 1 min and extension at 72°C for 1 min.

Rapid amplification of cDNA ends (5′-RACE)

Full-length pFS6 and pFS17 cDNAs were obtained using the 5′-AmpliFINDER™RACE kit (Clontech Laboratories) following the manufacturer's instructions. Poly(A)$^+$ RNA (2 μg) isolated from 13-DPA fibres was converted into first-strand cDNA using specific internal antisense primers 5′-GGC TTC CAC TAG CTG GTG AAC TTT CC-3′ and 5′-CAT TCG AAT GAG TAC AGC CAT GAG C-3′, designed from the pFS6 and pFS17 sequences, respectively. After anchor ligation PCR was carried out using a second pair of specific internal antisense primers (5′-GGG AAT TCC AAA ACT TCA CTT GAC GCT GTT GC-3′, for pFS6; 5′-GGG AAT TCG CAA CTG GTC TCT CTT CAA ACA TG-3′, for pFS17) and an anchor primer supplied by the manufacturer, together with Bresatec *Taq* Polymerase. PCR products were purified by centrifugation through a Sepharose CL-6B column, restricted with *Eco*RI and cloned into pBluescript®SK(−) (Stratagene) following routine protocols (Sambrook et al. 1989).

Southern blotting, Northern blotting and hybridisation

DNA samples were digested with restriction endonucleases under conditions recommended by the manufacturer (Boehringer Mannheim). Following electrophoresis, DNA was transferred to a Hybond-N$^+$ membrane (Amersham), following the manufacturer's instructions. Tissue specificity of isolated cDNA fragments was assayed by Northern blotting. Approximately 10-μg samples of total RNA were size-fractionated by electrophoresis on denaturing formaldehyde agarose gels (Sambrook et al. 1989) and transferred onto Nytran®-plus (Schleicher and Schuell) following the manufacturer's instructions.

DNA fragments requiring radioactive labelling were purified from an agarose gel using the freeze-squeeze method (Thuring et al. 1975). Probes were labelled with α-[^{32}P]dATP (Amersham) using the GIGAprime DNA labelling kit (Bresatec).

Prehybridisation, hybridisation and washes of Southern blots and Northern blots were carried out under conditions recommended by Amersham. Oligo(dA) (Amersham) to a concentration of 5 μg/ml was added to hybridisations involving cDNA library clones. Filters were exposed to X-ray film (Fuji RX) at −70°C in the presence of intensifying screens (DuPont Hi-Plus).

Sequencing of double-stranded DNA templates

pFS17 5′-RACE clones were sequenced using oligonucleotide primers designed to the insert sequence(s). Primers were synthesised on an Oligo100 DNA Synthesiser (Beckman) and purified using a DNA Ultrafast Cleavage and Deprotection kit (Beckman), following the manufacturer's instructions. Plasmid DNA for sequencing was purified by centrifugation through a Sepharose CL-6B chromatography column and sequenced using a Sequenase Version 2.0 kit (United States Biochemical Corp) and α-[^{32}P]dATP (Amersham). DNA sequencing reactions were electrophoresed in a 5% (w/v) acrylamide, 7 *M* urea and 1 × TBE (0.9 *M* TRIS-borate, 0.002 *M* EDTA) denaturing polyacrylamide gel using a BRL sequencing gel electrophoresis system. Sequence analysis utilised the DNA Strider software package and the GCG Sequence Analysis Software Package version 8 (Genetics Computer Group, Madison,

Wis. 1984). BLASTN and BLASTX were used to search nucleotide and protein sequence databases, respectively (Altschul et al. 1990; Gish and States 1993).

Results

An immature cotton fibre contains thousands of active genes (Graves and Stewart 1988) among which will be those that are crucial to fibre development. We have attempted to identify genes that are specifically expressed in 13-DPA fibres through a differential screening experiment. Screening of 2.3×10^4 recombinant complementary DNA (cDNA) clones with leaf and fibre cDNA probes resulted in the identification of 24 putative fibre-specific clones. Cross-hybridisation eliminated duplicate clones, reducing the putative fibre-specific clone population to six different sequence types (Table 1), and PCR across each insert revealed the relatively short average length of the inserted cDNAs. Hybridisation of individual clones to the cDNA library indicated that all correspond to abundant transcripts in 13-DPA fibre cells, ranging from 0.16% (pFS3) to 0.65% (pFS6) of the mRNA.

To investigate the spatial and temporal expression patterns of the cDNA clones, we carried out RNA blot analysis using RNA from various cotton tissues and from fibres of various developmental stages. Five of the six clones (pFS3, pFS5, pFS6, pFS17 and pFS18) showed highly preferential hybridisation to fibre mRNA (Table 1, Fig. 1), hybridising strongly to transcripts in fibre RNA but not to RNAs from leaf, whole flower or seedling tissue. One clone, pFS20, was unique in the clone collection and did not represent a fibre-specific transcript. Rather, pFS20 transcripts were present in fibre, leaf, flower and seedling tissues, producing constitutive signals comparable with those obtained for a cotton rDNA clone (not shown). This hybridisation pattern provides a useful positive control, reflecting RNA loading (Fig. 1A, B). On longer exposure, leaf, flower and seedling tissues showed weak hybridisation signals to each fibre-specific cDNA probe (not shown), indicating that transcription is not entirely restricted to the fibre or that products of other genes with related sequences cross hybridise to the cDNA probes.

The temporal pattern of transcript accumulation varied between clones, as shown by hybridisation of cDNA inserts to RNA from fibres of varying ages (Fig. 1, lanes 6–20). Genes corresponding to pFS5, pFS6 and pFS17 appeared to be expressed early in fibre development (6–14 DPA), whilst the pFS3 gene was expressed in fibres aged 6–20 DPA and expression of pFS18 appeared to peak at 12–14 DPA. The estimated sizes of the transcripts specified by the five fibre-specific clones ranged between 650 nt and 1500 nt, as indicated (Fig. 1), and the pFS20 cDNA hybridised to a 300-nt RNA species. Based on a comparison of cDNA length and transcript size from RNA blots, we concluded that full-length clones were not obtained for any of the fibre-specific mRNAs (Table 1).

Sequence information for cDNA clones corresponding to the five differentially-expressed mRNAs was used to determine potential functions of the encoded proteins. The sequence of pFS5 showed 96% nucleotide

Table 1 Putative fibre-specific cDNA clones from cotton

Representative cDNA clone	Number of clones in group	Transcript size (nt)[a]	cDNA size (bp)[b]
pFS3	1	700	519
pFS5	6	1000	509
pFS6	12	650	149
pFS17	3	1500	137
pFS18	1	700	97
pFS20	1	300	Not sequenced

[a] Estimated by comparison with RNA molecular weight markers (Promega)
[b] Determined by sequencing entire cDNA insert. Values do not include poly(A)$^+$ tail

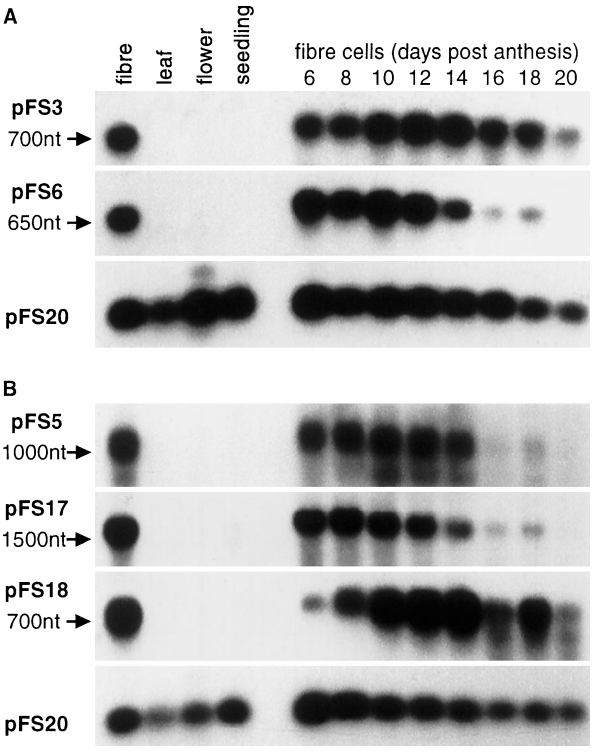

Fig. 1A, B Northern analysis of six putative fibre-specific cDNA clones. **A** and **B** represent independent Northern blots with 10 μg of total RNA in each lane, sequentially hybridised with the specific probes indicated. Total RNA loadings in each track were monitored by hybridisation with cDNA clone pFS20 and shown in the *lower panel* of each blot. Approximate transcript sizes in nucleotides (nt) are indicated on the *left*. The pFS18 hybridisation pattern was obtained using a longer homologue, identified by rescreening the cDNA library with pFS18

identity with the corresponding region of a previously characterised fibre-specific cDNA, E6 (John and Crow 1992). The 91 residue amino acid sequence derived from clone pFS5 is 91% similar to the 94 carboxy-terminal amino acids of the E6 protein (John and Crow 1992). A significant difference is an in-frame deletion of nine nucleotides from pFS5 compared with E6, which results in the absence of three amino acids from the conceptual translation product of pFS5. Sequences of pFS3, pFS6, pFS17 and pFS18 initially identified no definitive homologies in the nucleotide databases (GenBank and EMBL).

Further pFS6-related fibre-specific cDNA clones were sequenced, ranging in size from 131 bp (pFS24) to 347 bp (pFS19), and alignment of seven of the clones revealed variation in both sequence and length at their 3′ ends (Fig. 2A). Analysis of the 3′-untranslated regions of the cDNAs revealed the presence of a polyadenylation signal sequence AAUAAA, 89 nt upstream from the site of polyadenylation. In addition, there are two divergent polyadenylation signals, AAUAUU and AAUUAA, which occur 23 nt and 4 nt upstream from the site of polyadenylation respectively (Fig. 2A). An intermediate class of transcript includes pFS2, pFS21, pFS23 and pFS24, which are 6 bp longer than pFS6, the shortest cDNA. The remaining clones, pFS8 and pFS19, constitute a class of cDNA which is 23 bp longer than pFS6. Five nucleotide substitutions were observed in pFS24, together with a deletion of one nucleotide from the pFS6 sequence (Fig. 2A) compared with a consensus sequence, pFS6c.

The pFS6 consensus sequence of 329 bp shows similarity to cDNAs encoding phospholipid transfer proteins (LTPs), a class of plant proteins thought to be involved in the biogenesis of cellular membranes (Arondel and Kader 1990). In particular, the pFS6c nucleotide sequence is very similar (97%) to that of GH3, a cotton fibre-specific LTP cDNA recently isolated from *Gossypium hirsutum* L. cv 'DES119' (Ma et al. 1995). The conceptual translation of the pFS6c cDNA sequence shows 91% amino acid identity to the corresponding region of the GH3 protein, containing five amino acid substitutions of which only three are to amino acids with similar physical properties. Notably, the corresponding region of GH3 was identical to the sequence of pFS24, which was unique in the group of pFS6-like mRNAs sequenced in our experiments.

The remaining portion of the pFS6 mRNA was isolated *via* rapid amplification of cDNA ends (5′-RACE), using oligonucleotide primers complementary to the pFS6c cDNA sequence. Cloning of the resultant 500-bp 5′-RACE-PCR product produced two clones, pFS6-30 and pFS6-56, which were sequenced and, as expected, contained inserts which overlapped with pFS6c. The two 5′-RACE sequences were identical except for a C at position 104 in pFS6-30 which is substituted with a T in pFS6-56. In addition, pFS6-56 contained an attached, truncated PCR product of 111 bp which was identical

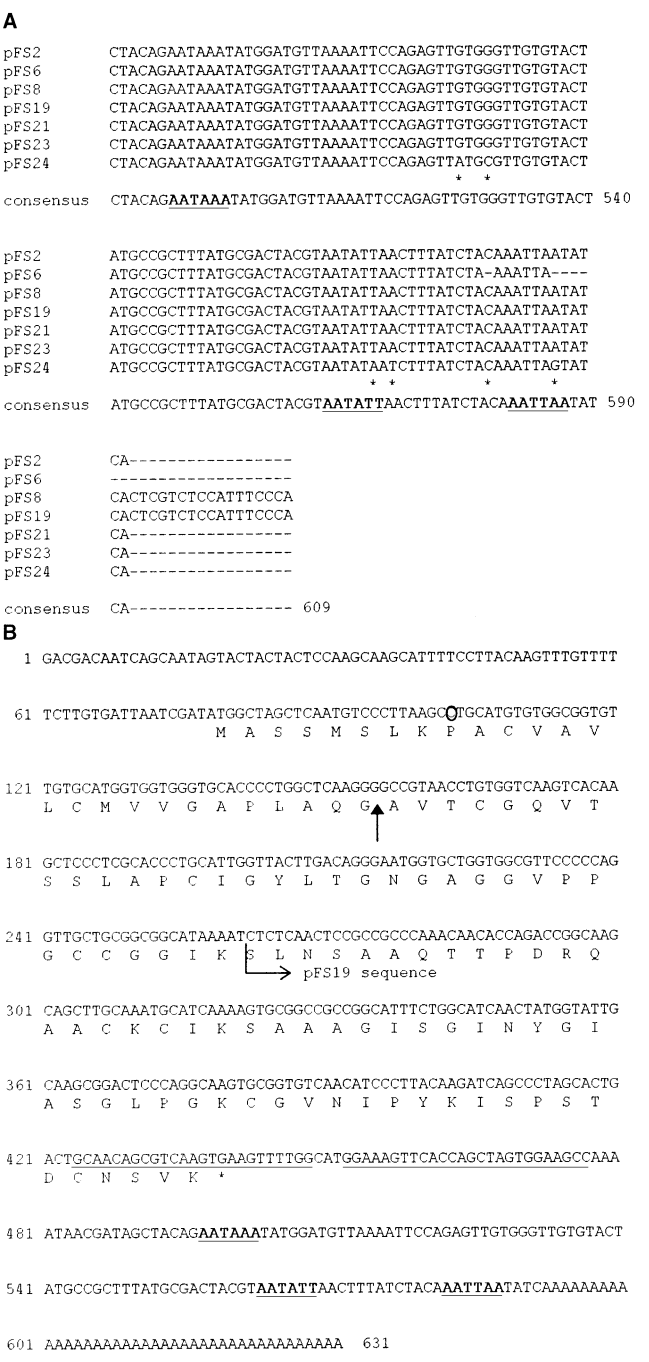

Fig. 2A, B Sequence of pFS6 group cDNA clones. **A** Alignment of the 3′-termini of cDNA clones which belong to the pFS6 group. Sequences are shown from nucleotide 491 of the full-length sequence (see Fig. 2B), and sequence differences are marked with an *asterisk*. Putative polyadenylation signal(s) are in *bold* and *underlined*. **B** Nucleotide sequence and deduced amino acid sequence of the full-length cotton LTP cDNA, cFSltp, constructed from the pFS6 cDNA consensus sequence and the sequence of the pFS6-30 clone. The putative signal peptide cleavage site is marked with an *arrow* and the pFS6 sequence indicated. The regions homologous to the oligonucleotide primers used for 5′-RACE are *underlined* and putative polyadenylation signal(s) are *underlined* and in *bold*. The nucleotide which differs between the two 5′-RACE clone sequences is *circled*

to the corresponding region of GH3 (Ma et al. 1995). The full-length cDNA, designated cFSltp, was derived from the sequences of clones pFS6c and pFS6-30 (Fig. 2B). The sequence of 592 bp contains a single significant open reading frame potentially encoding a 120-amino acid polypeptide, designated FSLTP. After the ATG translation initiation codon, there is a stretch of 26 amino acids which contains many of the features characteristic of signal peptides (Nothwehr and Gordon 1990). The putative cleavage site of the signal peptide was predicted according to von Heijne (1983) and confirmed by alignment with homologues from other plants (Fig. 3A).

The amino acid sequence of the mature FSLTP polypeptide shows only 88% identity with the conceptual translation of the previously characterised fibre-specific LTP cDNA, GH3 (Ma et al. 1995). In addition, the cotton fibre LTP sequence is similar to those found in spinach (Bernhard et al. 1991), tobacco (Fleming et al. 1992), tomato (Torres-Schumann et al. 1992), broccoli (Pyee et al. 1994) and carrot (Sterk et al. 1991) with which it shares amino acid identities of 77%, 64%, 63%, 57% and 54%, respectively (Fig. 3A). A hydropathy plot of FSLTP reveals a generally hydrophobic protein with few charged residues and a profile comparable to hydropathy plots of other plant LTPs (Fig. 3B). This pattern is very consistent with that of the tobacco and carrot homologues, to which the FSLTP protein is 64% and 54% similar, respectively. The other cotton LTP (GH3; Ma et al. 1995) is by far the most similar to cFSltp at the nucleotide level (95%) and amino acid level (88%) and yet appears quite different from FSLTP and the other LTPs in hydropathy profile (Fig. 3).

In a similar way, the 5′ end of the pFS17 mRNA was obtained using 5′-RACE. Cloning of a 1-kb 5′-RACE PCR product resulted in the isolation of three clones which were fully sequenced. Differences between the sequences (Fig. 4) can probably be attributed to *Taq* polymerase errors during 5′-RACE PCR since in the nine cases, one clone mismatched the other two by a single nucleotide substitution. A 1133-bp sequence for the full-length pFS17 transcript, designated cFS17, was constructed from the 5′-RACE consensus sequence and the pFS17 cDNA sequence (Fig. 4). Several open reading frames were identified in the cFS17 sequence, the largest of which encodes a putative polypeptide of 299 amino acids, designated FSPRP (Fig. 4) with a possible N-terminal signal peptide of 23 amino acids. The predicted signal sequence cleavage site was calculated according to von Heijne (1983), and processing would result in a mature protein of 276 amino acids with a calculated molecular weight of 31 567 Da and a pI of 10.5.

A comparison of the full-length pFS17 cDNA sequence with sequences in the databases (GenBank and EMBL) revealed significant similarity at both the nucleotide and amino acid levels to a class of structural

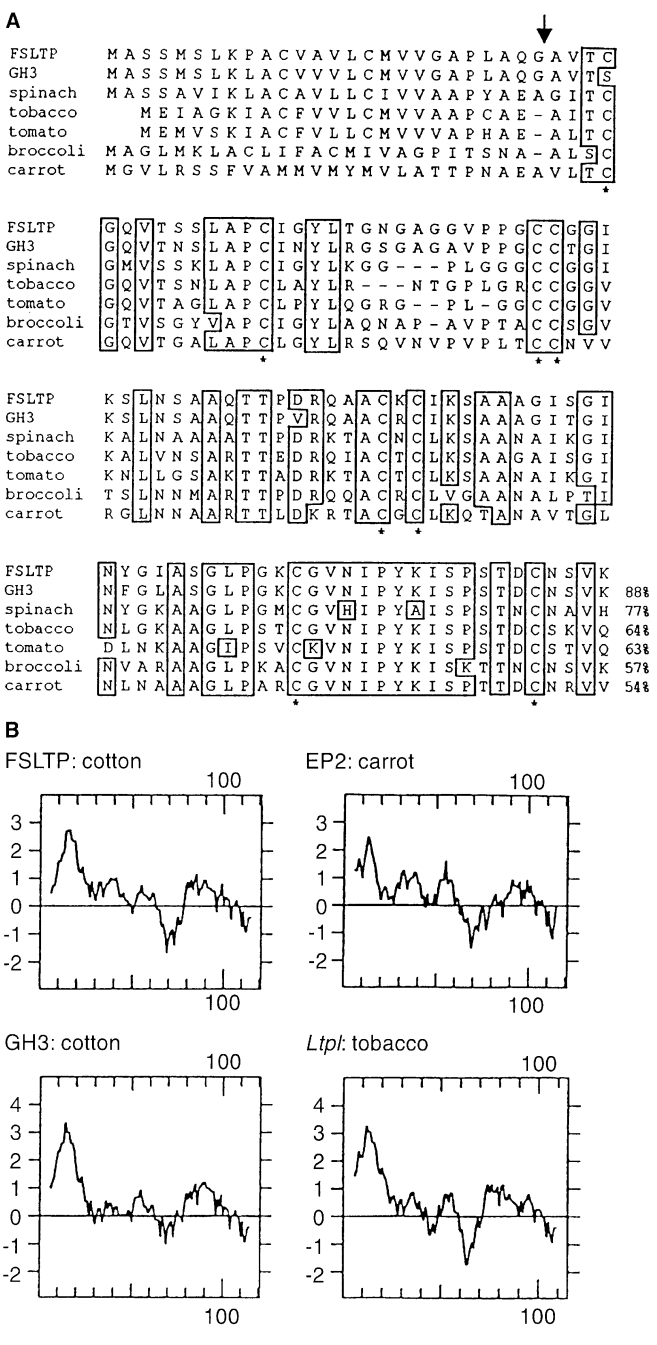

Fig. 3A, B Analysis of the 120 amino acid FSLTP protein. **A** Sequence comparison of seven plant lipid transfer proteins. *Boxes* indicate homologous amino acids in at least six of the sequences, and homology with *G. hirsutum* L. FSLTP (mature peptide only) is indicated at the end of each sequence. Putative signal peptide cleavage sites are marked with an *arrow*, and the highly conserved cysteine residues are indicated with *asterisks*. **B** Comparison of the hydropathy plots of two cotton fibre-specific LTPs with the homologues from carrot (EP2; Sterk et al. 1991) and tobacco (*Ltp1*; Fleming et al. 1992). The *horizontal scale* indicates the number of amino acid residues, and *the vertical one* the relative hydropathy scale (Kyte and Doolittle 1982). Points *above* the horizontal line correspond to hydrophobic region, and points *below* this line are hydrophilic

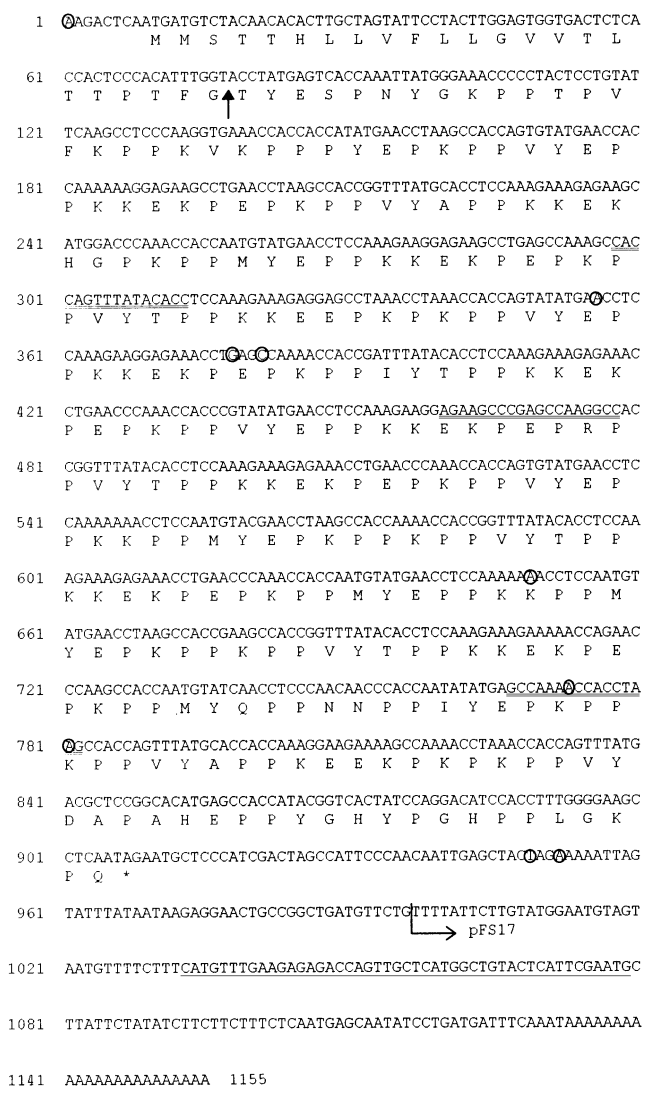

Fig. 4 Sequence of a cotton fibre-specific PRP cDNA. Nucleotide sequence and deduced amino acid sequence of the full-length cotton PRP cDNA, constructed from the pFS17 cDNA sequence and the consensus sequence of the pFS17 5′-RACE clones. The putative signal peptide cleavage site is marked with an *arrow* and the pFS17 sequence indicated. The regions complementary to the oligonucleotide primers used for 5′-RACE overlap by two nucleotides and are *underlined*, and regions complementary to primers used for sequencing the 5′-RACE clones are *double underlined*. Nucleotides which differed between the 5′-RACE clones are *circled*

cell-wall proteins, proline-rich proteins (PRPs) from several plants. The highest degree of sequence similarity was found to a soybean cDNA coding for a repetitive proline-rich protein SbPRP1 (Hong et al. 1987) with 57% nucleotide identity over a 385-bp overlap. At the protein level, the FSPRP and SbPRP1 sequences were 47% identical over a 154-amino acid overlap.

The cotton fibre PRP displays many of the features common to plant cell-wall PRPs. The repetitive nature of the FSPRP sequence is reflected in its hydropathy profile and is typical of plant PRPs but which can also diverge widely between PRPs from different plants (Fig. 5A). The PRP from soybean, SbPRP1 (Hong et al. 1987), contains 43 repeats of a sequence consisting primarily of PPVYK, whereas the less well-defined repetitive elements in a carrot PRP, DcPRP1 (Ebener et al. 1993), are interspersed with unique regions of sequence (Fig. 5A). The FSPRP protein can be divided into four regions based on its repeat unit structure. The first includes the hydrophobic signal peptide and the 28 N-terminal residues of the mature protein, both of which have little or no repeat sequence (Fig. 5B, 1). The centre region of the protein consists of two distinct parts. The first contains small repeats of 15 residues each, within which can be identified regions similar to the general element PPVYK (Fig. 5B, 2). The second repetitive region consists of three direct tandem repeats of 34 residues each (Fig. 5B, 3). The carboxy terminus of the FSPRP protein, like the amino terminus, is apparently devoid of repeat structure (Fig. 5B, 4).

The pFS18 clone was used to isolate additional cDNA clones of which two, of 616 and 610 bp, were analysed in detail (Olesnicky et al., unpublished results). These clones appear to contain only one significant open reading frame encoding 71 amino acids which showed strong similarity (88%) at the N-terminus to the transit peptide of pFS6 and other plant LTPs (result not shown). The remaining amino acids showed very low homology to pFS6, but three of the four conserved cysteines (Fig. 3A) are maintained. The large differences in temporal expression of pFS6 and pFS18 (Fig. 1) suggest that the proteins encoded by the two cDNA types are exported from the fibre cells at very different times in fibre development but that they utilise a very similar transit peptide. Other than the region encoding this transit peptide, these pFS18-like cDNAs showed no similarity to any other known sequences.

Discussion

Transcripts corresponding to each of five cDNA sequences accumulated to high levels in cotton fibre tissue but also exhibited low-level expression in leaf, whole flower and seedling tissue. This may be indicative of the presence of several multigene families in cotton, the different members of which are differentially expressed and/or encode divergent mRNA species. While it is unknown whether these mRNAs are translated, differing temporal expression patterns strongly suggest a role for transcriptional control of gene expression in cotton fibre cells.

Three of the five fibre-specific clones show substantial similarity to known proteins. The sequence of pFS5 showed similarity at the nucleic acid level and amino acid level to a previously characterised fibre-specific cDNA, E6 (John and Crow 1992). The sequence differences may be ascribable to allelic differences between

G. hirsutum L. cv 'Coker 312' and *G. hirsutum* L. cv 'Siokra 1-2', or to differences between homologues from the A and D subgenomes of the tetraploid *G. hirsutum* (for a review, see Endrizzi et al. 1985). The function of the E6 protein is unknown, but it has been suggested to have a unique structural or enzymatic role in cotton fibre primary cell-wall deposition. Predominance of pFS5 transcripts early in fibre development contrasts with the reported expression pattern of E6, which showed maximum mRNA levels in fibres aged 15–22 DPA (John and Crow 1992). However, early accumulation of pFS5 transcript in our experiments correlates very well with the immunodetection of the E6 protein product in Western analysis (John and Crow 1992).

The sequence of clone cFSltp, representing the most plentiful class of fibre-specific cDNA, showed significant similarity (50–90%) at both the nucleic acid and amino acid levels to lipid transfer proteins from a variety of plants. In particular, the eight cysteine residues are conserved (Fig. 2B), which makes it likely that the proteins fold into comparable three-dimensional structures (Tchang et al. 1988). The hydropathy profile of FSLTP is almost identical to those of tobacco and carrot but contrasts with GH3, another cotton fibre-specific LTP (Ma et al. 1995). The latter contains regions with reduced hydrophobic and hydrophilic domains. In common with other characterised LTPs, a putative 26- amino acid signal peptide sequence was identified within the FSLTP sequence, implying that the protein is secreted.

The nucleotide sequence of cFSltp shows 95% identity to that of GH3. The 120-amino acid conceptual translation of cFSltp shows only 88% amino acid identity to the GH3 protein, containing 14 amino acid changes, of which 10 are to amino acids with similar properties. One of the clones isolated in this study, pFS24, may be the GH3 homologue in 'Siokra 1-2', and it appears that a partial 5′-RACE product of 111 bp is an extension of pFS24. At least one class of LTP transcript is defined by the majority of the pFS6 homologues together with the 5′-RACE clones pFS6-30 and pFS6-56, while GH3, pFS24 and the truncated 5′-RACE product define a second class of fibre-specific

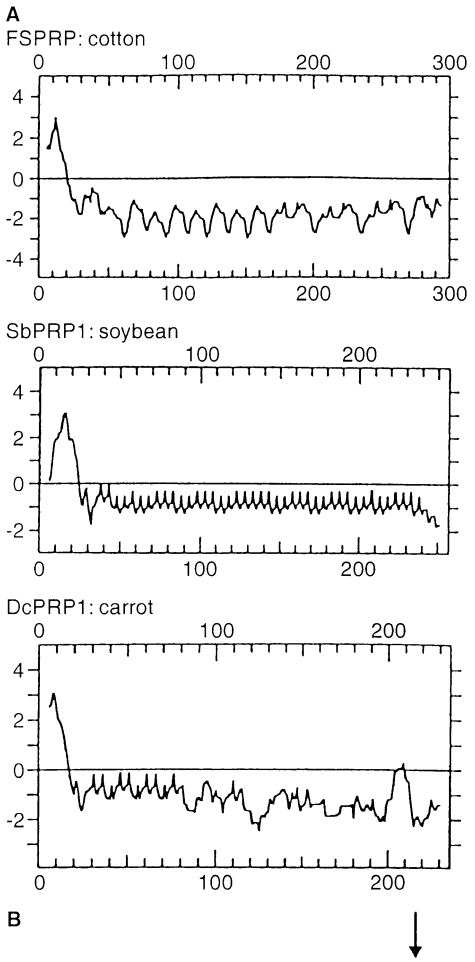

Fig. 5A, B Analysis of a fibre-specific PRP. **A** Comparison of the hydropathy profiles of a cotton fibre-specific PRP with homologues from soybean (SbPRP1; Hong et al. 1987) and carrot (DcPRP1; Ebener et al. 1993). The *horizontal scale* indicates the number of amino acid residues, and the *vertical one* the relative scale (Kyte and Doolittle 1982). Points *above* the horizontal line correspond to hydrophobic region, and points *below* this line are hydrophilic. **B** The amino acid sequence of a cotton fibre-specific PRP, divided into four regions or domains, based on its apparent repeat unit structure and the hydropathy profile. Sequences reminiscent of the general PRP element, PPVYK, are *underlined*, and domains are numbered *1–4*. *1* Amino acids 1–51 of the preFSPRP conceptual amino acid sequence, lacking in repeat structure. The putative signal peptide cleavage site is marked with an *arrow*. *2* Amino acids 52–176 of the preFSPRP conceptual amino acid sequence, arranged in tandem repeats of 15 amino acids each. Included is the intervening region between the blocks of repeats, consisting of 5 amino acids. *3* Amino acids 177–278 of the preFSPRP conceptual amino acid sequence, arranged in tandem repeats of 34 amino acids each. *4* Amino acids 279–299 of the preFSPRP conceptual amino acid sequence, lacking in repeat structure

LTP. Results from this study therefore suggest that cFSltp and GH3 are from different members of a fibre-specific LTP gene subfamily rather than the alternative that they reflect allelic variation.

The hypothesis that there is a fibre-specific LTP gene subfamily is supported by the isolation, in this experiment, of several fibre-specific LTP cDNAs which show variation near their 3′ termini. It may be that the different pFS6-like cDNA classes represent transcripts from distinct LTP gene family members of 'Siokra 1-2'. Another explanation is that the three cDNA length variants result from alternative use of the three putative polyadenylation sites within the cDNA sequences. Multiple polyadenylation sites are a common feature of plant genes, and substantial variation from the canonical signal, AAUAAA, can be tolerated (Wu et al. 1982).

LTPs are abundant proteins which may comprise up to 4% of the total soluble protein within a cell (Kader 1990). Northern analysis of plant LTP genes reveals complex expression patterns, which are often characterised by a high tissue specificity. For example, there are LTPs whose expression is restricted to the tapetal cells of the anther (Koltunow et al. 1990), to epidermal cells of somatic embryos (Sterk et al. 1991), to petal and carpel tissues (Kotilainen et al. 1994), or which are induced during seed germination (Linnestad et al. 1991; Torres-Schumann et al. 1992). Plant LTPs are generally encoded by multigene families of at least two non-allelic members, with different genes of the family encoding similar proteins that may have distinct expression patterns and/or functions.

Secretion of the protein has been demonstrated in the cases of spinach (Bernhard et al. 1991), barley (Mundy and Rogers 1986) and carrot (Sterk et al. 1991). LTPs isolated from both *Arabidopsis* and broccoli have been located to the cell wall and in particular, the waxy cuticle (Pyee et al. 1994; Thoma et al. 1993) where insoluble cutin may prevent water loss from epidermal cells and provide a mechanical barrier to pathogen ingress (Kolattukudy 1981). The induction of an analogous activity during cotton fibre development may be related to the extensive cutin deposition which occurs during the elongation phase of growth (Seagull 1993) and is consistent with an abundance of pFS6 transcripts early in fibre development. The protective cuticle is absent in other cotton cell types except epidermal cells, and LTP gene expression in epidermal cells may account for the low levels of LTP transcripts present in leaf, flower and seedling tissue.

The FSPRP sequence, derived from pFS17, displays many of the features common to PRPs, which are identified by the occurrence of Pro-Pro repeats that are contained within a variety of larger repeat units (Keller 1993; Showalter 1993). Analysis of the FSPRP amino acid sequence revealed an N-terminal region with several of the characteristics of eukaryote signal peptides (Nothwehr and Gordon 1990), and this suggests that the FSPRP mature protein is targeted to the secretory pathway in fibre cells and thus localised in the extracellular matrix. The FSPRP repeat unit appears to be related to the general plant PRP motif PPVYK (Josè and Puigdomènech 1993). PRPs are encoded by a small multigene family in several plant species, the members of which are tightly regulated and exhibit cell- and tissue-specific patterns of expression as well as developmental control (Keller 1993; Showalter 1993; Schmidt et al. 1994).

The precise function of PRPs is unknown, but a general role in the strengthening of plant cell walls is envisaged. PRPs are not solubilised by high salt (Datta et al. 1989) and are thought to be bound to the cell wall in complex structures. They may form a cross-linking network, as proposed for extensins (Fry 1986; Cooper et al. 1987), or they may interact with other components of the cell wall in associations mediated by the proline residues of PRPs (Williamson 1994). Such a cross-linking network would strengthen the cell wall against environmental stress and pathogen attack (Bradley et al. 1992). Proteins such as PRPs are expected to be abundant in rapidly elongating cells such as cotton fibres, since such growth requires a continuous supply of wall materials. An increase in mRNA levels of a PRP has been observed in cells of a barley *slender* mutant (Schünmann et al. 1994), which is partially characterised by epidermal cells which are both longer and narrower than those of the wild type. It is notable that FSPRP transcripts are abundant early in fibre development, during elongation, and that mRNA levels decreased at the onset of secondary wall synthesis when mainly cellulose is deposited (Fig. 1B).

The nucleotide sequence of cFS17 and the deduced amino acid sequence did not show significant similarity to a proline-rich protein previously isolated from cotton fibres, H6 (John and Keller 1995). The H6 gene, like pFS17, is expressed predominantly in cotton fibre cells, and the transcript is present during early primary cell-wall formation. Cotton protein H6 contains 17 copies of a repetitive pentameric motif of (Ala/Ser)(Thr/Ser) Pro Pro Pro which is clearly different from the FSPRP repeat unit(s) and places H6 in the arabinogalactan group of proteins (AGPs). Another fibre-specific protein of unknown function, B6 (John 1995), contains two proline-rich peptide regions, both of which are dissimilar to the repeats identified in the FSPRP sequence.

Screening of nucleotide databases was uninformative for two clones, pFS3 and pFS18. The use of pFS18 as a probe to a cDNA library resulted in the isolation of longer polyadenylated homologues which still do not identify with any previously published sequences. The abundance of pFS18 mRNA appeared to peak at 12–14 DPA, prior to the time of transition between the elongation (primary wall synthesis) and secondary wall deposition phases of fibre growth. It seems likely that the levels of at least some of the polypeptides involved in secondary wall synthesis may markedly increase

during this transition. The protein(s) encoded by pFS18 may therefore have a role in secondary wall deposition, such as the one proposed for a fibre-specific GTP-binding protein (Delmer et al. 1995). Transcripts of pFS3 were present during both the primary and secondary cell- wall synthesis stages, which may suggest a role for the corresponding protein in either elongation or secondary wall deposition in the cotton fibre.

Knowledge gained from this work may allow manipulation of gene expression to modify cotton fibre characteristics such as length, strength and fineness. In addition, isolation of the corresponding regulatory sequences will allow accurate tissue targeting of heterologous gene expression specifically to the fibres in transgenic plants. Future research efforts will aim to isolate the genes which encode the fibre-specific mRNAs described in this study.

Acknowledgements We are grateful to Prof. J. McD. Stewart for helpful discussions and to Cotton Seed Distributors Ltd, Wee Waa, NSW, Australia for providing cotton seeds.This work was generously supported by the Cotton Research and Development Corporation (CRDC), Narrabri, NSW, Australia.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arondel V, Kader J-C (1990) Lipid transfer in plants. Experentia 46: 579–585

Basra AS, Malik CP (1984) Development of the cotton fibre. Int Rev Cytol 89:65–112

Beasley CA (1979) Cellulose content in fibres of cottons which differ in their lint lengths and extent of fuzz. Physiol Plant 45:77–82

Bernhard WR, Thoma S, Botella J, Somerville CR (1991) Isolation of a cDNA clone for spinach lipid transfer protein and evidence that the protein is synthesised by the secretory pathway. Plant Physiol 95:164–170

Bradley DJ, Kjellbom P, Lamb CJ (1992) Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defence response. Cell 70:21–30

Colton RT (1991) Cotton fibre quality. Agfacts, New South Wales Dep Agric

Cooper JB, Chen JA, van Holst G-J, Varner JE (1987) Hydroxyproline-rich glycoproteins of plant cell walls. Trends Biochem Sci 12:24–27

Datta K, Schmidt A, Marcus A (1989) Characterisation of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. Plant Cell 1:945–952

Delmer DP, Pear JR, Andrawis A, Stalker DM (1995) Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibres. Mol Gen Genet 248:43–51

Dhindsa RS, Beasley CA, Ting IP (1975) Osmoregulation in cotton fibre: accumulation of potassium and malate during growth. Plant Physiol 56:394–398

Ebener W, Fowler TJ, Suzuki H, Shaver J, Tierney ML (1993) Expression of DcPRP1 is linked to carrot storage root formation and is induced by wounding and auxin treatment. Plant Physiol 101:259–265

Endrizzi JE, Turcotte, EL, Kohel RJ (1985) Genetics, cytology and evolution of *Gossypium*. Adv Genet 23:271–375

Fleming AJ, Mandel T, Hofmann S, Sterk P, de Vries SC, Kuhlemeier C (1992) Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. Plant J 2:855–862

Fry SC (1986) Cross-linking of matrix polymers in growing cell walls of angiosperms. Annu Rev Plant Physiol 37:165–186

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3:266–272

Graves DA, Stewart J McD (1988) Chronology of the differentiation of cotton (*Gossypium hirsutum* L.) fibre cells. Planta 175:254–258

Gupta VB, Moharir AV, Panda BC (1979) The characterisation of some Indian cotton varieties with the help of electron microscopy and X-ray diffraction and the correlation of their structure with their mechanical properties. In: Harrison PW (ed) Cotton in a competitive world. The Textile Institute, Manchester, UK, pp 83–105

Hong JC, Nagao RT, Key JL (1987) Characterisation and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. J Biol Chem 262:8367–8376

Jaswandala RT, Singh YD, Chinoy JJ (1977) Auxin metabolism in developing cotton hairs. J Exp Bot 28:1111–1116

John ME (1995) Characterisation of a cotton (*Gossypium hirsutum* L.) fibre mRNA (Fb-B6). Plant Physiol 107:1477–1478

John ME, Crow LJ (1992) Gene expression in cotton (*Gossypium hirsutum* L.) fibre: cloning of the mRNAs. Proc Natl Acad Sci USA 89:5769–5773

John ME, Keller G (1995) Characterisation of mRNA for a proline-rich protein of cotton fibre. Plant Physiol 108:669–676

Josè M, Puigdomènech P (1993) Structure and expression of genes coding for structural proteins of the plant cell wall. New Phytol 125:259–282

Kader J-C (1990) Intracellular transfer of phospholipids, galacto lipids, and fatty acids in plant cells. Subcell Biochem 16:69–111

Keller B (1993) Structural cell wall proteins. Plant Physiol 101: 1127–1130

Kolattukudy PE (1981) Structure, biosynthesis, and biodegradation of cutin and suberin. Annu Rev Plant Physiol 32:539–567

Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB (1990) Different temporal and spatial gene expression patterns occur during anther development. Plant Cell 2:1201–1224

Kotilainen M, Helariutta Y, Elomaa P, Paulin L, Teeri TH (1994) A corolla- and carpel-abundant, non-specific lipid transfer protein gene is expressed in the epidermis and parenchyma of *Gerbera hybrida* var 'Regina' (Compositae). Plant Mol Biol 26:971–978

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Linnestad C, Lönneborg A, Kalla R, Olsen O-A (1991) Promoter of a lipid transfer protein gene expressed in barley aleurone cells contains similar *myb* and *myc* recognition sites as the maize *Bz-McC* allele. Plant Physiol 97:841–843

Ma D-P, Tan H, Si Y, Creech RG, Jenkins JN (1995) Differential expression of a lipid transfer protein gene in cotton fibre. Biochim Biophys Acta 1257:81–84

Mundy J, Rogers JC (1986) Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: comparison to the barley amylase/subtilisin inhibitor. Planta 169:51–63

Naithani SC, Rama Rao N, Krishnan PN, Singh YD (1981) Changes in *o*-diphenol oxidase during fibre development in cotton. Ann Bot 48:379–385

Nothwehr SF, Gordon JI (1990) Targeting of proteins into the eukaryotic secretory pathway: signal peptide structure/function relationships. Bioessays 12:479–484

Pyee J, Yu H, Kolattukudy PE (1994) Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (*Brassica oleracea*) leaves. Arch Biochem Biophys 311:460–468

Ramsay JC, Berlin JD (1976) Ultrastructure of early stages of cotton fibre differentiation. Bot Gaz 137:11–19

Rowland SP, Nelson ML, Welch CM, Hebert JJ (1976) Cotton fibre morphology and textile performance properties. Text Res J 46: 194–214

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbour Laboratory Press, New York

Schmidt JS, Lindstrom JT, Vodkin LO (1994) Genetic length polymorphisms create size variation in proline-rich proteins of the cell wall. Plant J 6 : 177–186

Schubert AM, Benedict CR, Berlin JD, Kohel RJ (1973) Cotton fibre development-kinetics of cell elongation and secondary wall thickening. Crop Sci 13 : 704–709

Schünmann PHD, Harrison J, Ougham HJ (1994) *Slender* barley, an extension growth mutant. J Exp Bot 45 : 1753–1760

Seagull RW (1993) Cytoskeletal involvement in cotton fibre growth and development. Micron 24 : 643–660

Showalter AM (1993) Structure and function of plant cell wall proteins. Plant Cell 5 : 9–23

Sterk P, Booij H, Schellekens GA, Van Kammen A, de Vries SC (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. Plant Cell 3 : 907–921

Tchang F, This P, Stiefel V, Arondel V, Morch M-D, Pages M, Puigdomenech P, Grellet F, Delseny M, Bouillon P, Huet J-C, Guerbette F, Beauvais-Cante F, Duranton H, Pernollett J-C, Kader J-C (1988) Phospholipid transfer protein: full-length cDNA and amino acid sequence in maize. J Biol Chem 263 : 16849–16855

Thoma S, Kaneko Y, Somerville C (1993) A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. Plant J 3 : 427–436

Thuring RWJ, Sanders JPM, Borst P (1975) A freeze-squeeze method for recovering long DNA from agarose gels. Anal Biochem 66 : 213–220

Timpa JD, Ramey HH (1989) Molecular characterisation of three cotton varieties. Text Res J 59 : 661–664

Torres-Schumann S, Godoy JA, Pintor-Toro JA (1992) A probable lipid transfer protein gene in induced by NaCl in stems of tomato plants. Plant Mol Biol 18 : 749–757

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133 : 17–21

Wan C-Y, Wilkins TA (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal Biochem 223 : 7–12

Williamson MP (1994) The structure and function of proline-rich regions in proteins. Biochem J 297 : 249–260

Wu L, Ueda T, Messing J (1993) 3′-end processing of the maize 27-kDa zein mRNA. Plant J 4 : 535–544

Young RA (1986) Structure, swelling and bonding of cellulose fibres. In: Young RA, Rowell RM (eds) Cellulose, structure, modification and hydrolysis. John Wiley and Sons, New York, pp 91–128