PROGRESS IN THE IDENTIFICATION OF GENES WHICH CONTROL COTTON FIBRE YIELD AND QUALITY

Sharon J Orford, Sven K Delaney, John A Humphries,
Damien J Lightfoot and Jeremy N Timmis
School of Molecular and Biomedical Science, The University of Adelaide, Adelaide SA, 5005

Introduction

Cotton fibres, or lint, are very long single cells containing almost pure cellulose. The fibres develop in the weeks after flowering from single cells on the surface of the young seed. Each fibre cell is small at first but elongates and develops rapidly, eventually forming the mature cotton fibre. These processes require the ordered expression of genes which make enzymes, structural proteins and signaling molecules that together determine the properties of the fibre.

Domestication has modified fibre development to produce cotton varieties with greatly improved fibre length, strength and quality. However, the selection and breeding of plants with desirable fibre characteristics is slow and expensive. As such, future crop improvement is likely to depend upon genetic engineering and the cotton industry has been a leader in both research and commercialisation of transgene technology, with momentous consequences for the agronomic properties of the crop such as insect and herbicide resistance. However, the potential for fibre modification by biotechnology has yet to be realised.

Apart from their commercial significance, the single-celled nature of cotton fibres and their synchronous growth inside the cotton boll has made them an attractive system for the study of plant cell elongation and cell wall biogenesis. Such fundamental studies have resulted in the identification of a number of genes with proposed roles in cotton fibre growth, and have led to the creation of the first molecular and cellular model of fibre development (Wilkins and Jernstedt, 1999). The model has provided the basis for limited early success in the alteration of fibre growth, but not on a commercial scale.

An expansin gene expressed exclusively in cotton fibres

Previous CRDC-funded research in Adelaide has focused on identifying genes that are expressed at high levels in fibre cells but not in other tissues, since these genes are likely to control fibre characteristics such as yield and quality. All six genes identified in our laboratory (Orford and Timmis, 1997; Orford and Timmis, 1998) are structural genes which encode proteins mostly involved with cell wall synthesis, a result which is not surprising given that cell wall deposition is the primary activity of the rapidly growing cotton fibre.

The most interesting isolate encodes an expansin (Orford and Timmis, 1998). Expansins are a large family of proteins which loosen the components of rigid plant cell walls, thereby allowing a cell to expand. Six members of the expansin gene family in cotton were characterised in detail (Harmer *et al.*, 2002), including one which is expressed at high levels during elongation of cotton fibre cells. This expansin may be pivotal to fibre cell initiation and elongation and is therefore an attractive target in the design of strategies to alter fibre growth.

To this end, we made four genetic constructs, in which expansin gene expression in the fibres is prolonged, down-regulated or up-regulated. The promoters were available from our bank of fibre-specific gene promoters (see next section). The gene constructs were used to transform whole cotton plants (at CSIRO Division of Plant Industry in Canberra) and a large number of transformed lines have been recovered. T1 plants have been grown for at least one line per experiment and tested for presence of the transgene, with T2 seed being collected from positive plants. T2 plants will be tested for the effects of the transgene on fibre properties such as length, strength, micronaire, uniformity and maturity. These experiments provide the first example of temporal alteration of the expression of a native cotton gene, and offer exciting prospects for the genetic improvement of cotton fibre length.

Dissection of cotton fibre-specific promoters

The expression of genes is controlled by regions known as promoters, which determine the timing, location and level of gene expression in response to signals from inside and outside the cell. We have isolated the promoters which correspond to six different fibre-expressed genes (see above), and these provide important tools in the genetic engineering of cotton fibres in several respects. Firstly, expression of transgenes only in the fibre cells avoids any potential problems from gene expression in non-fibre tissues. Secondly, they regulate the level and timing of transgene expression and so control the effects of transgenes on fibre characteristics. Finally, the promoters may be used with any transgene, and could therefore be used to produce, for example, naturally pigmented cotton fibres or fibres with enhanced insect resistance. In addition, desired changes in fibre characteristics could be achieved by manipulating gene expression by altering the promoters themselves.

The aim of this part of the project is to determine how the promoters direct the timing, level and location of gene activity, and to identify the signals to which they respond. Sequence analysis of our six fibre-specific promoters has identified a number of elements, such as TATA and CAAT boxes (necessary for basal gene expression), elements that may respond to MYB transcription factors and elements that mediate responses to plant hormones. However, the precise sequence elements that regulate fibre-specific gene expression cannot yet be identified by computer comparisons, and we are using an experimental approach in order to detect such elements.

Firstly, we tested our candidate fibre-specific promoters for their ability to direct expression of a reporter gene in transient assays in cotton fibres and other tissues by microprojectile bombardment. Gene constructs were made in which each of the six promoters was fused to a reporter gene, βglucuronidase (GUS). All six promoters were shown to be fibre-specific. That is, the reporter protein was detected only in the fibres and not in bombarded leaf or petal tissue. The same set of constructs was then used to transform whole cotton plants (being carried out at CSIRO Division of Plant Industry in Canberra) and 50 different plant lines (at least one per experiment) have been recovered. T1 plants were grown and tested by PCR for presence of the transgene, with T2 seed being collected from positive plants. Quantitative GUS assays will be carried out on tissues collected from T2 transgenic plants in order to verify the fibre-specificity and determine the strength of each promoter.

One of the fibre-specific promoters, pPR6, which is thought to be the most active in fibres, has been analysed in detail. Tobacco and *Arabidopsis* trichomes (hairs) are an attractive model for cotton fibre growth because of the developmental similarities between cotton fibres and trichomes, and the cost and time factors involved in transforming tobacco and *Arabidopsis* (approx 3 months) as opposed to cotton (at least 12 months). Transgenic tobacco and *Arabidopsis* plants with stable genomic integration of the pPR6::GUS reporter construct have therefore been generated. A striking result was obtained, in which GUS expression was detected only in the leaf trichomes of transgenic tobacco and in the trichomes and leaf margins of transgenic *Arabidopsis*.

Identification of fibre-specific promoter elements (FSEs) in the pPR6 promoter has been approached by a strategy in which progressive 5' deletions are tested for their ability to drive fibre-specific GUS expression. A series of constructs was assembled and delivered to cotton fibres, leaves, petals and bracts by microprojectile bombardment. An 84 bp region of the promoter was found to be necessary for GUS expression exclusively in the cotton fibres and represents the "minimal" fibre-specific promoter. A yeast monohybrid assay was used to isolate the cotton fibre proteins that interact with this promoter region, and sequence analysis demonstrated that the interactors can be divided into three categories on the basis of homology with known plant transcription factors. Future work will involve isolation of the full-length cDNAs of these proteins and characterisation of their function using green fluorescent protein (GFP)-based cellular localisation studies and electromobility shift assays (EMSA). This information will be used to generate improved and novel fibre-specific expression systems with potential commercial application.

Gene expression in the cotton fruit

The *Cry* genes from *Bacillus thuringiensis* have been engineered into a number of commercially grown cotton varieties, including INGARD® and BollgardII® cotton in Australia, and their introduction has had very positive impacts on the industry in terms of economic benefit to growers, increased insect control and reduction in pesticide usage. In these transgenic crops, the *Bt* toxin is expressed throughout all tissues of the plant, at a high metabolic expense to the cotton plant. This constitutive expression (~1% of total soluble protein) is not necessary for Australian cotton crops because of the fruit-directed feeding habits of the key cotton pests (*Helicoverpa sp.*). It would be advantageous to restrict expression of the toxin to the cotton boll wall, thereby forming a physical barrier of toxin-expressing tissue to protect the fruit and fibres from pest attack. An additional benefit of reduced *Bt* toxin manufacture is that mortality among non-target species would be lessened and the other tissues of the plant, such as leaves and seeds, would be free of *Bt* toxin (or other transgene) such that, for example, seeds would be more acceptable for oil production or animal feed. Notably, the outcomes of this work would not be limited to use with the *Bt* transgenes and could be readily used for any other transgene.

Current research is therefore aimed at the isolation of genes (and their promoters) which are expressed only in the outer wall of the cotton boll. The differential screening technique, which proved successful in the isolation of fibre-specific genes, was applied to a 5 DPA boll wall cDNA library. A large number of putative boll wall-specific mRNAs were identified and their expression patterns determined by Northern analysis. Three cDNAs had appropriate expression patterns (i.e. high and specific boll wall expression) and sequencing revealed similarity to proline rich proteins, chalcone synthase and starch synthase. Genomic Southern blots showed that each is a member of only a small gene family, and a PCR-based genome walking technique will be performed to isolate the corresponding promoter regions. The promoter sequences will be initially analysed by computer-based methods to find common or known motifs, and transient bombardment assays will be performed to verify the boll-specific expression patterns of the promoters. Such work will result in the development of novel promoters, capable of driving boll wall-specific expression of a transgene (or transgenes) in commercial cotton cultivars.

The molecular switch to fibre cell differentiation

An intriguing aspect of cotton fibre growth is the process by which fibre cells are determined. All ovule epidermal cells are able to differentiate into fibre cells, but only about one in four elongate to become fibres. Although the anatomy of young fibres has been well-studied by microscopy, nothing is known about the molecular basis of cotton fibre initiation and the distribution of fibre cells on the seed, except that the process appears to be controlled by plant growth factors such as auxins and gibberellins (Beasley and Ting, 1974). A second type of differentiated epidermal cell, "fuzz", initiates growth up to several days after the lint fibres and, being very short, has little or no economic value. The timing of initiation of cotton fibres is therefore an important step in obtaining a commercially useful product.

A number of investigators including ourselves have proposed that the trichomes of *Arabidopsis thaliana* are a suitable model system for determining the genetic basis of cotton fibre initiation. *Arabidopsis* trichomes are single-celled hair-like structures which are found on all parts of the plant, and share a number of developmental similarities with cotton fibres. The simplicity, visibility and dispensability of trichomes in *Arabidopsis* make them an ideal system for studying plant cell differentiation and morphogenesis, with the result that the three or four genes required for normal trichome initiation have been fully characterised, and a model of the control of *Arabidopsis* trichome differentiation is emerging (Szymanski *et al.*, 2000).

Mutations in only two genes, both encoding transcription factors, result in a complete lack of trichomes. One of the genes, *TRANSPARENT TESTA GLABRA1* (*AtTTG1*), is required for both trichome initiation and spacing in *Arabidopsis* and encodes a protein of the WD-repeat protein family, members of which have been implicated in a diverse range of cellular processes such as signal transduction, RNA processing, gene regulation, vesicular traffic and regulation of the cell cycle (Walker *et al.*, 1999; Yu *et al.*, 2000). Using a combination of PCR-based and library screening techniques, we have isolated and characterised four cotton homologues of *AtTTG1*.

The four cotton sequences fall into two distinct pairs, with one pair (*GhTTG1* and *GhTTG3*) having 80% amino acid identity to *AtTTG1*, and the other pair (*GhTTG2* and *GhTTG4*) around 62%. All four genes are derived from the ancestral D diploid genome of tetraploid cotton and are expressed in many tissues throughout the plant, including ovules and growing fibres. Functional testing of the genes was approached by stable transformation of *ttg1* mutant *Arabidopsis* plants, which have a number of defects including an absence of trichomes and seed coat pigments. Interestingly, both *GhTTG1* and *GhTTG3* were able to complement the *ttg1* mutation and restore normal trichome formation. The same two genes also complemented anthocyanin production in a white-flowered *Matthiola incana ttg1* mutant in which the cotton sequences were transiently expressed. These significant results indicate that two cotton genes are able to behave *in planta* like *Arabidopsis TTG1*, and may therefore be functional homologues of *TTG1*. It is tempting to speculate that such WD-repeat proteins could have a pivotal role in cotton fibre initiation. Our results demonstrate a molecular parallel between differentiation of trichomes in cotton and *Arabidopsis*, and provide the first functional characterisation of a regulator of cotton fibre cell growth (Humphries *et al.*, 2004). We are currently conducting a yeast two-hybrid screen in order to identify molecular interactors of the GhTTG3 protein.

The way forward

Since 1992 we have been conducting CRDC-funded research on the genes which are expressed during cotton fibre development. Several genes have been analysed in detail such that the work has recently entered a new and exciting phase. We are now in a position to carry out functional analysis of developmental genes and cotton fibre-specific promoters, and to test the effects of perturbing the expression of a particular gene on cotton fibre characters. Cotton transformation experiments are coming to fruition, with a large number of transformant lines available for analysis.

Findings from our research will contribute substantially to our knowledge of the molecular control mechanisms which underlie fibre cell initiation, development and agronomic properties. More generally, the project will contribute to our knowledge of cell fate determination and control of gene expression in plants by transcription factors. The research will significantly advance gene discovery in cotton, particularly key genes that control growth and development of the cotton fibre, and hence, determine economically important properties. Such information will provide a clear basis for fibre quality improvement in Australian cotton cultivars through genetic engineering, an outcome which can only result in economic benefits for the Australian cotton industry.

Acknowledgements

S Orford is grateful for the continued support for this research by the Cotton Research and Development Corporation, Narrabri, New South Wales. S Delaney, J Humphries and D Lightfoot are all generously supported by Postgraduate Scholarships from the Cotton Research and Development Corporation. These experiments comply with the current Australian statutory requirements

References

Beasley CA and Ting IP (1974) Effects of plant growth substances on *in vitro* fibre development from unfertilised cotton ovules. *Amer J Bot* 61: 188-194

Harmer SE, Orford SJ and Timmis JN (2002) Characterisation of six α-expansin genes in *Gossypium hirsutum* (upland cotton). *Mol Genet Genomics* 268: 1-9

Humphries JA, Walker AR, Timmis JN and Orford SJ (2004) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana TRANSPARENT TESTA GLABRA1* (*TTG1*) gene. *Submitted for publication*

Orford SJ and Timmis JN (1997) Abundant mRNAs specific to the developing cotton fibre. *Theor App Genet* 94: 909-918

Orford SJ and Timmis JN (1998) Specific expression of an expansin gene during elongation of cotton fibres. *Biochim Biophys Acta* 1398: 342-346

Szymanski D, Lloyd A and Marks M (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. *Trends Plant Sci* 5: 214-219.

Walker A, Davison P, Bolognesi-Winfield A, James C, Srinivasan N, Blundell T, Esch J, Marks M and Gray J (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11: 1337-1349.

Wilkins TA and Jernstedt JA (1999) Molecular genetics of developing cotton fibres. In Basra AS (ed) *Cotton fibres: developmental biology, quality improvement, and textile processing*. The Haworth Press, Inc, NY, pp 231-269.

Yu L, Gaitatzes C, Neer E and Smith TF (2000) Thirty-plus functional families from a single motif. *Prot Sci* 9: 2470-2476.