MOLECULAR CLONING AND ANALYSIS OF THE GENE FOR A COTTON FATTY ACID DESATURASE (FAD2)

I. L. Pirtle, W. Kongcharoensuntorn, M. Nampaisansuk, K. D. Chapman and R. M. Pirtle
University of North Texas
Denton, TX
J. E. Knesek
Texas Woman's University
Denton, TX

Abstract

Over 75% of fatty acids in plants are unsaturated by desaturase enzymes in chloroplast and endoplasmic reticulum membranes. Two fatty acid desaturases designated FAD2 and FAD3 primarily desaturate extrachloroplast lipids and occur as integral membrane proteins in the endoplasmic reticulum. The *FAD2* gene has been shown by several laboratories to be important in the chilling sensitivity of plants, since polyunsaturated phospholipids are essential for maintaining plant viability at lowered temperatures. Cotton genomic libraries harbored in lambda phage were screened with an *Arabidopsis FAD2* hybridization probe to isolate prospective *FAD2* genes, in order to study their regulation of gene expression. Two overlapping genomic clones were found to encompass a *FAD2* gene by alkaline blot hybridization and DNA sequence analysis. Since restriction fragments from the cloned DNAs correspond to identically-sized cotton genomic DNA fragments, the cloned DNA fragments represent actual genomic fragments encompassing the *FAD2* gene. The protein-coding region of this gene is 1,155 bp, and is continuous with no introns. The deduced amino acid sequence of 384 amino acids of the putative cotton FAD2 polypeptide has a high identity (about 75%) with the deduced amino acid sequences of other plant FAD2 enzymes, such as *Arabidopsis* FAD2 and soybean FAD2. The cotton FAD2 enzyme has histidine-rich motifs that could serve as potential iron-binding domains for electron transport for the desaturation reaction, similar to other plant FAD2 amino acid sequences. Yeast cells transformed with a plasmid construct containing the cotton *FAD2* coding region have an appreciable accumulation of linoleic acid (18:2), not normally present in wild-type yeast cells. Thus, this cotton *FAD2* gene is truly functional, since it encodes an enzyme that catalyses the desaturation of oleic acid (18:1) into linoleic acid (18:2). The *FAD2* gene has one large intron of 2,967 bp in the 5'-flanking region, only 11 bp upstream from the ATG initiation codon. The presence of a large intron in the 5'-flanking region could be important in the transcriptional regulation of this gene. The *FAD2* gene promoter/enhancer motif has a potential TATA basal promoter element and two presumptive basic region-helix-loop-helix (bHLH) motifs with the consensus sequence CANNTG. The bHLH or E box motif has been implicated as a seed-specific regulatory element.

Introduction

Fatty acids in plants, as in all other organisms, are the major structural components of biological membranes (phospholipid bilayers) and storage oils (neutral lipids or triacylglycerols). Thus, it is important to understand the mechanisms underlying the regulation of fatty acid compositions in membrane phospholipids in plants. A major control point may be at the level of gene expression, such as how the genes involved in lipid biosynthesis are regulated in plants and if there are *trans*-acting transcription factors simultaneously controlling the expression of many genes in lipid synthesis. The major site of the 30 enzymatic reactions to produce 16- and 18-carbon fatty acids in plants is the stroma of plastids, with over 75% of fatty acids being unsaturated (Ohlrogge and Browse, 1995). Desaturation of the fatty acid components of the membrane phospholipids is carried out by membrane-bound desaturases of the chloroplast and endoplasmic reticulum membranes. The *FAD2* and *FAD3* enzymes primarily desaturate extrachloroplast lipids and occur as integral membrane proteins in the endoplasmic reticulum. The enzymes are phosphatidylcholine desaturases, acting on fatty acids at both the sn-1 and sn-2 positions. The *FAD2* gene appears to be especially important in the chilling sensitivity of plants (Miquel and Browse, 1994; Okuley et al., 1994; Ohlrogge and Browse, 1995). Thus, polyunsaturated phospholipids in membranes are essential for maintaining cellular function and plant viability at lowered temperatures. In cotton, a knowledge of the tissue-specific and development-specific regulation of fatty acid compositions of the membranes is crucial in understanding the heat and cold tolerance, desiccation and salt tolerance, and disease resistance. A clearer understanding of the expression patterns of the *FAD2* gene(s) will allow us to genetically engineer this gene in cotton, and in turn, ultimately manipulate the fatty acid compositions of cotton membranes predictably to improve the vigor and viability of this important crop plant.

Materials and Methods

Isolation of Genomic Clones

To prepare hybridization probe for screening a cotton genomic library to isolate clones encompassing presumptive *FAD2* genes, a 0.5-kb SalI fragment was purified from an *Arabidopsis* cDNA clone designated FAD2-43 (GenBank T13887) generously provided by Dr. T. Newman of the *Arabidopsis* Biological Resource Center, Ohio State University (Newman et al., 1994; Okuley et al., 1994). This DNA fragment was used as a template to generate ^{32}P-labeled DNA fragments by the random priming procedure (Feinberg and Vogelstein, 1983) for use in screening one cotton (*Gossypium hirsutum*, cv. Acala SJ-5) genomic library, harbored in the vector λEMBL3 (Grula et al., 1995), kindly provided by Dr. David M. Anderson of Phytogen Seeds, Placentia, CA. Plaque hybridization was done at 55°C by the procedure of Benton and Davis (1977). Several plaques that gave intense, positive signals were selected for plaque purification. The cloned DNAs were initially analyzed by a mini-lysate procedure involving restriction endonuclease digestion, agarose gel electrophoresis, and alkaline blot hybridization. Prehybridization of the positively charged nylon membrane (Amersham Hybond N$^+$) filter replicas was done in a solution consisting of 6 X SSC, 0.5% SDS, 5 X Denhardt's reagent, and denatured sheared salmon sperm DNA (100 µg/ml) for 4 hours. Hybridization was done in a solution of 6 x SSC, 0.5% SDS, 5 X Denhardt's reagent, 20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 2.5 mM Na pyrophosphate (pH 8.0), denatured sheared salmon sperm DNA (100 µg/ml), and [^{32}P]-labeled probe for approximately 18 hours. After hybridization, the nylon membranes were washed once for five min. and once for thirty min. in 2 x SSC, 0.1% SDS, and then twice for 30 min. in 1 X SSC, 0.1% SDS.

Based on the tentative physical map, a 10-kb HindIII fragment of a genomic clone designated LCFg55 containing a putative *FAD2* gene was subcloned into the HindIII site of the cosmid vector pDELTA2 (Gibco BRL) for DNA sequence analysis. Since the clone LCFg55 was found to not contain a full-length *FAD2* gene, it became necessary to isolate an overlapping genomic clone. A second cotton (*Gossypium hirsutum*, cv. Acala SJ-2) genomic library harbored in the Lambda FIXII vector (Stratagene), generously donated by Dr. Thea Wilkins (Univ. of California at Davis), was then screened to isolate additional clones encompassing the entire *FAD2* gene. A unique, *FAD2* gene-specific 5'-flanking region fragment was generated by PCR amplification of 390 bp of the partial 5'-flanking region of the *FAD2* gene in LCFg55, and inserted in the vector pGEM7Zf(+) to prepare ^{32}P-labeled hybridization probe by random priming. Conditions used to isolate the genomic clone LCFg24 were the same as described above for LCFg55, except the hybridization temperature was increased to 60°C.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:551-554 (2001)
National Cotton Council, Memphis TN

Physical Mapping and Genomic Blot Hybridization

To construct the physical maps of the cotton genomic inserts in the lambda clones, LCFg24 and LCFg55 DNAs were digested with restriction endonucleases and the fragments analyzed by agarose gel electrophoresis and alkaline blot hybridization (Reed and Mann, 1985). Either a 0.8-kb XhoI/EcoRI fragment from a corresponding cotton *FAD2* cDNA clone designated pSKCF106A or a 0.5-kb SalI fragment from the *Arabidopsis* cDNA clone designated FAD2-43 (Okuley et al., 1994) were used for template to generate hybridization probe by random priming.

Cotton genomic DNA, extracted from young cotton leaves (cv. Paymaster HS26) by the method of Paterson et al. (1993), was digested with the restriction endonucleases HpaI and HindIII, and doubly-digested with HindIII and SalI. Cloned LCFg24 DNA was digested with the same enzymes. The fragments were resolved on a 0.8% agarose gel, alkaline-blotted, and hybridized with ^{32}P-labeled DNA probe generated by random priming from a homologous 1.2-kb cotton DNA template from the *FAD2* coding region.

DNA Sequence Analysis

Both strands of three plasmid subclones of the two overlapping lambda clones LCFg55 and LCFg24 encompassing the *FAD2* gene were sequenced using a primer-based approach. Initially, a 10.0-kb HindIII fragment of LCFg55 was inserted into the cosmid vector pDELTA2 for transposon-based intramolecular generation of nested deletion subclones for DNA sequencing (Gibco BRL Life Technologies Deletion Factory 2.0 Manual). Overlapping deletion subclones were generated, and their DNA sequences analyzed on a semi-automated LI-COR 4000L instrument using the dideoxynucleotide-chain termination method with the Epicentre SequiTherm Excel II Long-Read cycle sequencing system. The non-overlapping regions of the cotton DNA segment were sequenced (or the sequences of some GC-rich areas with compression effects confirmed) using manual sequencing with the chain termination method with a thermostable DNA polymerase (*Thermo*Sequenase from Amersham) and ^{33}P-labeled dideoxynucleotides for terminator cycle sequencing (Fan et al., 1996). In addition, a 6.8-kb HindIII fragment and a 4.1-kb SalI fragment of the clone LCFg24 were subcloned into the plasmid pUC19 for both semi-automated and manual sequencing. To generate nested random deletions of the LCFg24 subclones, the EZ::TN Plasmid-Based Deletion Machine method (Epicentre) was used. Analysis of the DNA sequence was done with DNASIS software (Hitachi). The transcription start point (prospective cap site) and the locations of the 5'- and 3'-splice sites of the 5'-flanking intron were deduced from a comparison of the gene sequence with the sequence of the 5'-untranslated region of a PCR-amplified *FAD2* cDNA.

Construction of a Yeast Plasmid Vector for Functional Expression of the *FAD2* Coding Region

To construct a recombinant plasmid in the 5.6-kb yeast shuttle expression vector pYES2 (Invitrogen), the cotton *FAD2* coding region in a plasmid subclone of LCFg55 was amplified by PCR, using a forward primer with a SacI site and a reverse primer with an EcoRI site. The 1.5-kb PCR product was directionally subcloned into the SacI/EcoRI polylinker site of pYES2 in the sense orientation, downstream from the galactose-inducible *GAL1* promoter. The 7.1-kb construct designated pYES2/FAD2-3 was electroporated into both yeast INVSc1 cells (Invitrogen) and *E. coli* DH5α cells. The DNA sequence of the *FAD2* insert was confirmed to be identical with that of the *FAD2* coding region in LCFg24 and LCFg55. In addition, the pYES2 vector was also electroporated into yeast INVSc1 cells for use as a control.

Both yeast transformants were grown in SC-U medium (synthetic complete minus uracil) (Adams et al., 1998) at 30°C overnight, washed, and suspended in galactose induction medium (SC-U medium containing 2% galactose and 2% raffinose) and grown for three generations. The cells were pelleted and washed four times with water to remove any media or metabolites which could interfere with the lipid analyses. The fatty acids were extracted and transmethylated with 5% HCl in methanol at 85°C for three hours. The fatty acid methyl esters were analyzed by gas chromatography and quantified by flame ionization detection (FID) essentially as described by Chapman and Trelease (1991).

Results and Discussion

The cotton genomic inserts in LCFg24 and LCFg55 were deduced to be 12.3 kb and 13.9 kb, respectively, by physical mapping. Since the two genomic fragments overlap by 9.6 kb, an overall 16.5-kb region of cotton genomic DNA is harbored in the two clones. Initially, we isolated and characterized the *FAD2* gene in the genomic clone LCFg55. The *FAD2* gene in LCFg55 lacks the 5'-flanking promoter/enhancer region and most of the 5'-flanking intron, as determined by DNA sequence analysis. A second cotton genomic library was then screened, and a lambda clone designated LCFg24 was characterized by physical mapping and DNA sequence analysis. LCFg24 was found to contain the 5'-flanking intron and promoter regions of the *FAD2* gene. The 7,914-bp genomic sequence with the *FAD2* gene has been assigned GenBank Number AF331163. The cotton DNA segment in the clone LCFg24 must represent actual cotton genomic DNA, since 6.8-kb HindIII, 2.9-kb HindIII/SalI, and 2.6-kb HpaI fragments from cloned LCFg24 DNA correspond to identically-sized cotton genomic fragments. In addition, the HindIII and HindIII/SalI digests of genomic DNA have additional hybridizing fragments that must be derived from another *FAD2* gene. Thus, there are two (or more) actual *FAD2* genes in the cotton genome, which would likely be the case, since cotton is allotetraploid.

The coding region of the *FAD2* gene has 1,155 basepairs, including the termination codon, and is continuous with no introns. The deduced amino acid sequence of the FAD2 polypeptide is 384 amino acids. Since the 2,967-bp intron is spliced out to generate the mature *FAD2* mRNA, the 5'-flanking untranslated region (5'-UTR) of the mature mRNA is 130 nucleotides. The 5'-flanking intron occurs only 11 bp upstream from the ATG initiation codon. The *Arabidopsis FAD2* gene (Okuley et al., 1994) also has a rather large, 1,100-basepair intron in its 5'-flanking region, only four basepairs from its ATG initiation codon. The presence of large introns in the 5'-flanking regions of *FAD2* genes could be important in the transcriptional regulation of expression of these genes. The 5'-flanking region of the *FAD2* gene has several potential promoter/enhancer elements that could function as positive regulatory elements in gene expression. A TATA basal promoter element occurs 41 bp upstream from the putative cap site. Two basic region helix-loop-helix (bHLH) or E box motifs with the consensus sequence CANNTG occur at 109 bp and 135 bp upstream from the potential cap site. The E box motif has been shown to be a seed-specific regulatory element in the French bean β-phaseolin gene (Kawagoe et al., 1994).

The gene has a 3'-untranslated region (3'-UTR) of 242 basepairs, deduced by comparison with the 3'-flanking region of the corresponding cotton *FAD2* cDNA clone pSKCFc106A (GenBank Number AF329635) sequenced in our laboratory. The near-upstream polyadenylation signal (AATCAA) occurs 30 basepairs upstream from the 3'-polyadenylation site in the putative *FAD2* mRNA transcript.

There is 85% identity between the predicted amino acid sequences of the cotton FAD2 in this report and the cotton FAD2-2 (Liu et al., 1997; GenBank Y10112) and 74% identity between the cotton FAD2 and the cotton FAD2-1 (Liu et al., 1999; GenBank X97016). Similar identities of about 75% occur between the cotton FAD2 and the soybean, *Arabidopsis*, and *Borago* FAD2 proteins. The cotton *FAD2-1* cDNA clone encodes a seed-specific transcript that is specifically induced in embryo development and that is not detected in vegetative tissues like leaves, most likely having

a major role in converting oleate into linoleate during storage lipid biosynthesis in cottonseed (Liu et al., 1999). In contrast, the cotton *FAD2-2* gene (Liu et al., 1997) apparently has a low level constitutive expression in leaf tissue and throughout seed development, similar to the constitutive soybean *FAD2-2* gene (Hitz et al., 1994; Heppard et al., 1996). The *Arabidopsis FAD2* gene (Okuley et al., 1994) has been shown to be equally expressed in both vegetative and embryo tissues. Since the cotton FAD2 protein (this report) has a higher identity to the cotton FAD2-2 protein than the cotton FAD2-1 protein (Liu et al., 1997; 1999), it may be expressed constitutively throughout the cotton plant rather than being involved in seed-specific gene expression.

Acyl-lipid and acyl-CoA desaturases are hydrophobic proteins that apparently span the membrane four times (Los and Murata, 1998; Shanklin et al., 1994). The enzymes in the endoplasmic reticulum of plant cells accept electrons from an electron transport system composed of cytochrome b_5 and NADH-dependent cytochrome b_5 reductase (reviewed in Los and Murata, 1998; Shanklin and Cahoon, 1998). A hydropathy plot of the cotton FAD2 polypeptide was generated by the method of Kyte and Doolittle in DNASIS software (Hitachi), and there are six presumptive membrane-spanning domains that correspond with the predicted membrane-spanning domains in the desaturase integral membrane protein models of Los and Murata (1998) and Shanklin et al. (1994). In addition, all known desaturases are characterized by the presence of three histidine clusters localized at very conserved locations in the protein sequence (Los and Murata, 1998; Shanklin and Cahoon, 1998). These three histidine clusters are thought to comprise the catalytic center of the enzyme, since they may form ligands to a diiron cluster in the catalytic site (Shanklin and Cahoon, 1998). The amino acid sequence of the putative FAD2 polypeptide has the three histidine-rich motifs (Los and Murata, 1998; Shanklin and Cahoon, 1998) in the conserved locations in the protein structure.

Yeast cells cannot convert oleic acid (18:1) into linoleic acid (18:2), since they do not have *FAD2* genes. For this reason, the transformation of yeast cells with the cotton *FAD2* gene should lead to synthesis of linoleic acid from endogenous oleic acid present in yeast cells. This conversion would lead to a dramatic shift in the fatty acid profile of yeast cells harboring the *FAD2* gene, since there should be a substantial increase in linoleic acid and a commensurate relative decrease in the amount of oleic acid. Yeast cells transformed with the plasmid construct pYES2/FAD2 were found to have an appreciable accumulation of linoleic acid (18:2), not normally present in wild-type yeast cells. A slight amount of palmitoleic acid (16:2) was also observed. The 18:2 and 16:2 were not detected in the control yeast cells transformed by the vector pYES2. Also, the oleic acid (18:1) peak in the transformed cells was noticeably smaller than the corresponding 18:1 peak in the control cells, indicating the conversion of 18:1 into 18:2 in the yeast cells containing the cotton *FAD2* gene. Thus, it can be concluded that this cotton *FAD2* gene is truly functional, since it encodes an enzyme that catalyzes the desaturation of oleic acid.

Acknowledgements

This research was supported by Cotton Incorporated Agreement 99-652 overseen by Dr. Gay M. Jividen, Cary, NC, USDA Award 98-35503-6339 from the NRI Competitive Grants Program, Texas Advanced Technology Program Grant 003594-014, and University of North Texas Organized Research Funds.

References

Adams, A., D.E. Gottschling, C.A. Kaiser, and T. Stearns. 1998. Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., p. 148.

Benton, W.D. and R.W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196: 180-182.

Chapman, K.D. and R.N. Trelease. 1991. Acquisition of membrane lipids by differentiating glyoxysomes: role of lipid bodies. J. Cell Biol. 115: 995-1007.

Fan, J., R.S. Ranu, C. Smith, C. Ruan, and C.W. Fuller. 1996. DNA sequencing with [α-^{33}P]-labeled ddNTP terminators: a new approach to DNA sequencing with Thermo Sequenase DNA polymerase. BioTechniques 21: 1132-1137.

Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Guthrie, C., and G.R. Fink, eds. 1991. Guide to Yeast Genetics and Molecular Biology. In Methods in Enzymology, Vol. 194. Academic Press, New York, NY.

Grula, J.W., R.L. Hudspeth, S.L. Hobbs, and D.M. Anderson. 1995. Organization, inheritance and expression of acetohydroxyacid synthase genes in the cotton allotetraploid *Gossypsium hirsutum*. Plant Mol. Biol. 28: 837-846.

Heppard, E.P., A.J. Kinney, K.L. Stecca, and G.-H. Miao. 1996. Developmental and growth temperature regulation of two different microsomal ω-6 desaturase genes in soybeans. Plant Physiol. 110: 311-319.

Hitz, W.D., T.J. Carlson, J.R. Booth, Jr., A.J. Kinney, K.L. Stecca, and N.S. Yadav. 1994. Cloning of a higher-plant plastid ω-6 fatty acid desaturase cDNA and its expression in a cyanobacterium. Plant Physiol. 105: 635-641.

Kawagoe, Y., B.R. Campbell, and N. Murai. 1994. Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein beta-phaseolin gene. Plant J. 5: 885-890.

Liu, Q., S.P. Singh, C.L. Brubaker, P.J. Sharp, A.G. Green, and D.R. Marshall. 1997. In Physiology, Biochemistry and Molecular Biology of Plant Lipids, J. P. Williams, M.U. Khan and N.W. Lem, eds. Kluwer Academic Publishers, Boston, MA, pp. 383-385.

Liu, Q., S.P. Singh, C.L. Brubaker, P.J. Sharp, A.G. Green, and D.R. Marshall. 1999. Molecular cloning and expression of a cDNA encoding a microsomal ω-6 fatty acid desaturase from cotton (*Gossypium hirsutum*). Aust. J. Plant Physiol. 26: 101-106.

Los, D.A. and N. Murata. 1998. Review: Structure and expression of fatty acid desaturases. Biochim. Biophys. Acta 1394: 3-15.

Miquel, M.F. and J.A. Browse. 1994. High-oleate oilseeds fail to develop at low temperature. Plant Physiol. 106: 421-427.

Newman, T., F.J. de Bruijn, P. Green, K. Keegstra, H. Kende, L. McIntosh, J. Ohlrogge, N. Raikhel, S. Somerville, M. Thomashow, E. Retzel, and C. Somerville. 1994. Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol. 106: 1241-1255.

Ohlrogge, J. and J. Browse. 1995. Lipid biosynthesis. Plant Cell 7: 957-970.

Okuley, J., J. Lightner, K. Feldmann, N. Yadav, E. Lark, and J. Browse. 1994. Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6: 147-158.

Paterson, A.H., C.L. Brubaker, and J.F. Wendel. 1993. A rapid method for extraction of cotton (*Gossypsium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol. Biol. Rep. 11: 122-127.

Reed, K.C. and D.A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. 13: 7207-7221.

Shanklin, J., and E.B. Cahoon. 1998. Desaturation and related modifications of fatty acids. Annu. Rev. Plant Physiol. 49: 611-641.

Shanklin, J., E. Whittle, and B.G. Fox. 1994. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry 33: 12787-12794.