Cloning of a cDNA encoding diacylglycerol acyltransferase from *Arabidopsis thaliana* and its functional expression

Douglas H. Hobbs, Chaofu Lu, Matthew J. Hills*

John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

Received 22 April 1999

Abstract Triacylglycerols are the most important storage lipids in most plants and animals. Acyl-CoA:diacylglycerol acyltransferase (EC 2.3.1.20) catalyzes the final step of the pathway of triacylglycerol synthesis and is the only step which is unique to this process. Diacylglycerol acyltransferase is required for the synthesis of storage oil in a wide range of oil-bearing seeds and fruits and in floral structures such as petals, anthers and pollen. We describe the first cloning and functional expression of a cDNA encoding diacylglycerol acyltransferase from a plant. The cDNA, cloned from *Arabidopsis thaliana*, encodes a 520 amino acid protein with a predicted molecular mass of 59.0 kDa which shares 38% amino acid sequence identity with diacylglycerol acyltransferase from mouse. When expressed in insect cell cultures, the protein catalyzes the synthesis of [^{14}C]triacylglycerol from [^{14}C]diacylglycerol and acyl-CoA. Primer extension analysis revealed that the transcription begins 225 bases before the translation start site, yielding an unusually long 5′ untranslated region. The gene is expressed in a wide range of tissues but most strongly in developing embryos and petals of flowers.

Key words: Diacylglycerol acyltransferase; Triacylglycerol; Seed oil; *Arabidopsis thaliana*

1. Introduction

Triacylglycerols (TAGs) are stored by plants in tissues such as seeds, fruits, petals and anthers and are an important agricultural commodity. In many plants, TAGs are essential for at least part of the life cycle, e.g. mutants in seed oil synthesis in *Arabidopsis* are unable to germinate in the absence of a supplementary carbon source such as sucrose [1]. TAGs may also be important in the sexual reproduction of many plants since they can comprise 30–40% of the total pollen mass. The role of TAGs in pollen is not entirely clear but they may supply fatty acids for the rapid synthesis of membrane phospholipids to support pollen tube growth [2]. Recently, it has been shown that TAGs may also be required for the proper growth of the pollen tube in plant species that have wet stigmas such as tobacco [3].

TAGs are synthesized by the enzymes of the Kennedy pathway, which sequentially transfer acyl chains from acyl-CoAs to the *sn*-1, -2 and -3 positions of a glycerol backbone [4]. Diacylglycerol (DAG) lies at the branch point between membrane glycerolipid and storage TAG synthesis. Diacylglycerol acyltransferase (DGAT) catalyzes the final acylation of the pathway and is the only step which is unique to TAG synthesis. DGAT is a membrane bound enzyme that is thought to be located in the endoplasmic reticulum [5,6]. There are indications that DGAT might be an important step in the control of TAG synthesis [7] and cloning of the DGAT gene will facilitate investigation of its role(s). An *Arabidopsis* mutant in the TAG synthesis was shown to accumulate DAG in the seeds [8] but attempts to clone the gene involved have not yet been successful. Recently, a cDNA encoding a DGAT from mouse was cloned [9] and, using the sequence information of the mouse DGAT, we have cloned a cDNA and identified the corresponding gene encoding a plant DGAT homologue.

2. Materials and methods

2.1. Isolation of a cDNA clone encoding the DGAT enzyme

A BLAST search of the sequence databases using the mouse DGAT protein sequence [9] identified *Arabidopsis* EST (E6B2T7). The EST (Arabidopsis Stock Centre, OH, USA) was fully sequenced in both directions. The 5′ end of the cDNA was obtained by 5′ RACE with appropriate nested primers using a 5′ RACE PCR kit supplied by Gibco BRL (UK). Total RNA was isolated from tissues of *Arabidopsis thaliana* and *Brassica napus* [10]. First-strand cDNA was prepared from total RNA from *Arabidopsis* inflorescences, using oligonucleotide primer DGAT 2 (5′-d(AACCAAAGGTGGAAGAAGCAG)-3′). The resulting single-stranded cDNA was tailed at the 5′ end using dCTP catalyzed by TdT and the 5′ end of the DGAT cDNA was amplified by two rounds of PCR according to the manufacturer's instructions using nested primers DGAT 4 (5′-d(AGGATGCTTTGAGTTCCTGAC)-3′, reverse) and DGAT 5 (5′-d(AATCCGGTGAATATGACCAG)-3′). The DNA product (1200 bp) was cloned into pGEM T easy (Promega, UK) according to the manufacturer's instructions. The DNA inserts were completely sequenced in both directions using a Perkin Elmer Big Dye sequencing kit (Perkin Elmer, UK). Sequences were analyzed locally using the Wisconsin GCG programs. Database searches were done using the BLAST algorithm [11] and protein analysis was done using the Prosite database (http://expasy.hcuge.ch/cgi-bin/scanprosite).

2.2. Primer extension analysis

The transcription start site of the *DGAT* gene was localized by primer extension analysis using two gene-specific 30-mer oligonucleotides which were complementary to the sense strand sequence of the *DGAT* cDNA. Oligonucleotides PE1 (5′-CGCCATTTCGAAAAGCGTTTGACGAAACAG-3′), from −24 to +6 relative to the translation start site (ATG), and PE2 (from −141 to −112, 5′-GCGGATTGAAGGAAGAAGCTAAGAAATCGG-3′) were radiolabelled at their 5′ termini with T4 polynucleotide kinase and ([γ-^{32}P]ATP. About 10^5 cpm of the radiolabelled primers was hybridized at 30°C with ∼50 μg of total RNA, which was isolated from *Arabidopsis* flowers. After hybridization for 12 h, complementary DNA was synthesized from the annealed primers by the addition of reverse transcriptase and dNTP. The reaction was stopped by incubating at 70°C for 10 min and then treated with the RNAse mix. The nucleic acid was precipitated with ethanol. The reaction products were resus-

*Corresponding author. Fax: (44) (1603) 259882.
E-mail: hillsm@bbsrc.ac.uk

PII: S0014-5793(99)00646-8

pended in sequencing gel loading buffer, denatured at 85°C, resolved by electrophoresis through a 6% w/v polyacrylamide sequencing gel containing 7 M urea and visualized by autoradiography. In order to provide size markers, part of the *DGAT* gene was sequenced with the same primers used in the primer extension experiments.

2.3. Expression of DGAT in Sf21 insect cell cultures

A full-length DGAT cDNA was cloned by RT-PCR from *Arabidopsis* inflorescence RNA using the Gibco BRL 5′ RACE kit as described above. The reverse transcriptase reaction was primed using DGAT 15 (5′-d(TTGCAGATACAACAACTC)-3′) and the PCR reaction used the primers DGAT 15 and DGAT 13 (5′-d(CTATAAGGAACCAGACGC)-3′) with Pfu DNA polymerase. The 1800 bp product was cloned into pGEM T easy as described above to create plasmid AtflDGAT. The DGAT encoding sequence was obtained by PCR using primers DGAT 28 (5′-d(GTACGGATCCAAATGGCGATTTTGGATTC)-3′) and DGAT 29 (5′-d(GTACGATATCTTTCATGACATAGATCC)-3′). The product was sequenced in both directions, checked to be correct and cloned into pFASTBAC1 (Gibco BRL, UK). DH10Bac *Escherichia coli* cells were transformed with the product to generate recombinant bacmid DNA according to the manufacturer's instructions. Sf21 insect cells were transfected with the bacmid DNA and recombinant baculovirus particles harvested. The virus stock was amplified and the amplified viruses used to infect Sf21 cells.

2.4. Infection of insect cells, isolation of cell membranes and in vivo labelling of cell proteins and gel analysis

Insect cells (2.5×10^6 per 25 cm^2 dish) were infected with virus and at 48 h, the cells were removed, washed twice with phosphate-buffered saline and pelleted by centrifugation at $1000\times g$ for 5 min. The cells were homogenized and membranes purified as previously described [9]. For in vivo labelling, at 48 h post-infection, 53 MBq of Promix ([^{35}S]methionine and [^{35}S]cysteine: Amersham, UK) was added to the cell cultures. The cells were incubated for a further 2 h and the medium was discarded. Protein from about 3×10^5 cells was dissolved in gel sample buffer and resolved by SDS-PAGE on 10% polyacrylamide gels using a Bio-Rad mini-gel apparatus according to the manufacturer's instructions. Radiolabelled proteins were identified by autoradiography

2.5. Assay for DGAT activity

DGAT was assayed by incorporation of [1-^{14}C]dioleoylglycerol into TAG in the presence of oleoyl-CoA. Radiolabelled dioleoylglycerol was synthesized by phospholipase C hydrolysis of [1-^{14}C]dioleoyl phosphatidylcholine (110 mCi/mmol) and [^{14}C]DAG was purified by thin layer chromatography (TLC) as described in other methods. DGAT activity was assayed using a modification of a previously described protocol [9]. The reaction mixture (100 μl) contained 100 mM Tris-HCl, pH 7.5, 250 mM sucrose, 3 mM $MgCl_2$, 1 mM EDTA and 10 μM [^{14}C]DAG (3.92 kBq/nmol) in 0.02% Tween-20. Oleoyl-CoA was included at 50 μM where indicated. Reactions were initiated by the addition of 25 μg membrane protein and incubated at 30°C for 30 min. Incorporation of [^{14}C]DAG into TAG was determined as previously described [6]. Protein was estimated using a Coomassie dye-binding assay from Bio-Rad according to the manufacturer's instructions using thyroglobulin to generate the standard curve.

2.6. Northern blot analysis of RNA

20 μg of total RNA from tissues of *B. napus* was run on 1.2% agarose gels and transferred to charged nylon filters (Amersham Pharmacia, UK) as previously described. Blots were hybridized with ^{32}P-labelled DNA probes as previously described [12]. Radioactive probes were labelled using a kit from Stratagene according to the manufacturer's instructions. Filters were washed twice in 2×SSC, 0.1% SDS at 65°C for 15 min and twice with 0.2×SSC, 0.1% SDS at 65°C for 15 min. Hybridization was detected by autoradiography using a Kodak ARX film.

2.7. Lipid methods

Lipids were extracted from insect cells as previously described [6,13]. Total lipid extracts were applied to 250 μm silica layers (Anachem, UK), separated using hexane/diethyl ether/acetic acid (70:30:1, v/v/v) and visualized by charring the TLC plate with H_2SO_4 in ethanol.

3. Results and discussion

3.1. Cloning and analysis of cDNA encoding DGAT

A BLAST search of the sequence databases using the amino acid sequence of the mouse DGAT [9] showed that it shared a significant amino acid sequence homology with an *Arabidopsis* EST (E6B2T7). The EST was fully sequenced in both directions and analysis of the deduced amino acid sequence suggested that over 1000 bp of additional 5′ DNA were required to obtain a full-length open reading frame. DNA of the 5′ end was obtained by 5′ RACE from *Arabidopsis* inflorescence RNA using appropriate nested primers and was fully sequenced in both directions. The full-length cDNA is 1988 bp long, containing an open reading frame of 1563 bp encoding a protein of 520 amino acids (EMBL accession number AJ131831). The amino acid sequence shares 38% identity and 59% similarity with the mouse DGAT but only 24% identity with the mouse ACAT. The three sequences were aligned using the Pileup program (Fig. 1a) revealing regions of significant homology between the *Arabidopsis* protein and both mouse DGAT and ACAT (residues 287–309 and 370–398). There are a number of other regions of the *Arabidopsis* protein that show significant homology only to the mouse DGAT. The protein is predicted to have a molecular mass of 59.0 kDa with an iso-electric point at pH 8.53. The *Arabidopsis* protein is estimated to be 6.0 kDa larger than the mouse protein with most of the extra mass accounted for by a 38 amino acid N-terminal extension. Searches of the protein databases suggest that this extension is not involved in protein targeting.

A Kyte and Doolittle hydrophobicity plot suggests that the protein contains a number of membrane spanning domains but also a 14 kDa hydrophilic domain at the N-terminus (Fig. 1b). The amino acid sequence was also analyzed using a program which predicts transmembrane helices by comparison with a database of naturally occurring membrane bound proteins ([14]: http://ulrec3.unil.ch/software/TMPRED_form.html). This strongly predicts nine transmembrane helices with the large N-terminal domain lying on the cytoplasmic side of the membrane. Scanning the protein sequence against the Prosite database (http://expasy.hcuge.ch/cgi-bin/scanprosite) identified a number of putative phosphorylation sites and a leucine zipper between residues L-222 and L-250. It remains to be determined whether these are important in the regulation and function of the enzyme in vivo. No other motifs with a recognizable function were found. A Cys_6His_2 involved in the binding of 1,2-DAG in protein kinase Cδ and RasGRP [15] is not present in either the *Arabidopsis* or mouse sequences, suggesting that DAG is bound by a different structure in the acyltransferases. Comparison of the plant glycerol-3-phosphate acyltransferase and lysophosphatidatic acid acyltransferase sequences with the plant DGAT did not show any significant sequence homology, suggesting they are unrelated.

3.2. Expression of Arabidopsis DGAT in insect cells

The putative DGAT coding sequence was inserted into the pET15b vector for expression in *E. coli* without a His_6 affinity tag and into the pYES 2.0 vector for expression in *Saccharomyces cerevisiae*. Neither protein expression or enzyme activity could be detected in extracts from induced cultures of either system. Therefore, the coding sequence was sub-cloned

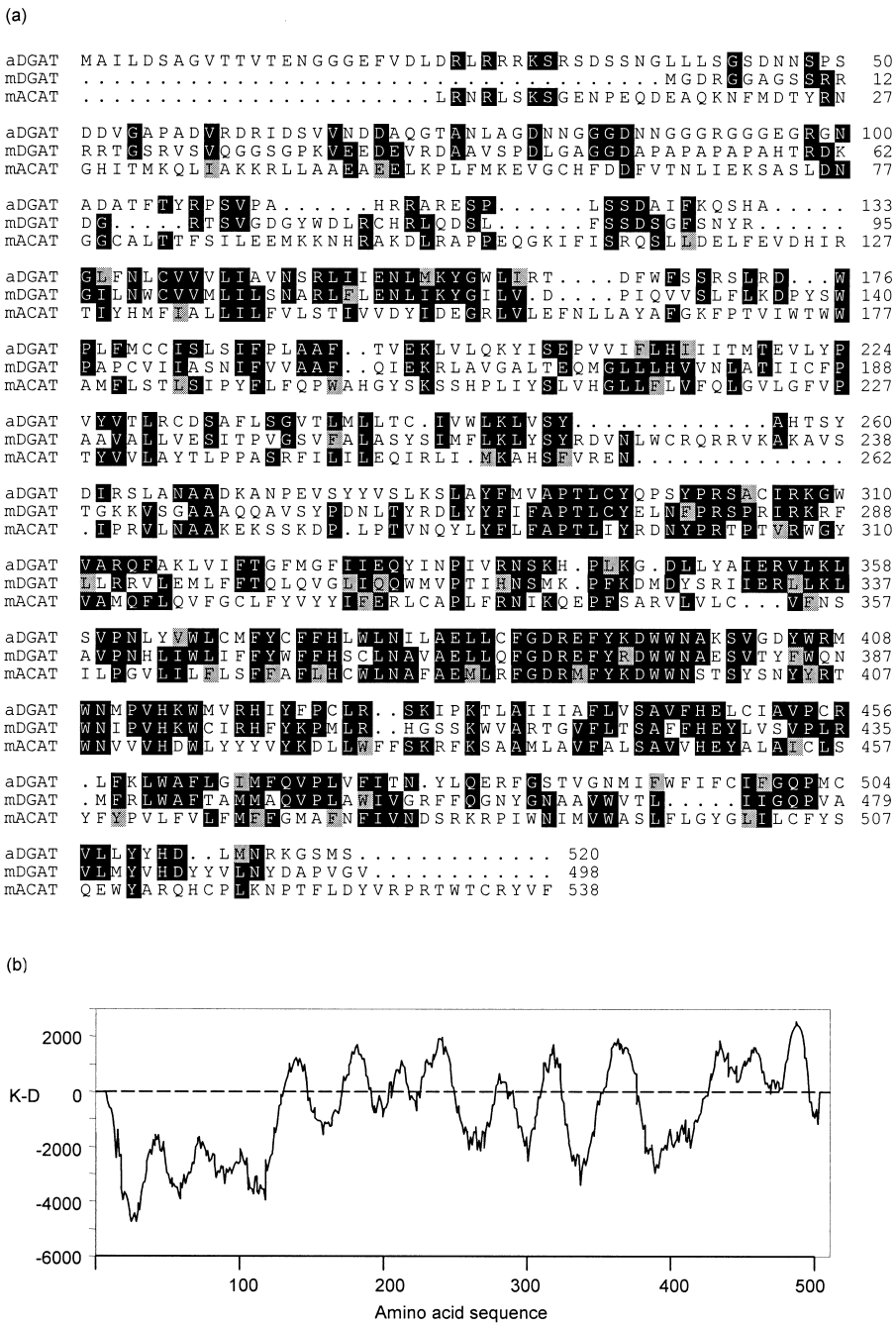

Fig. 1. Analysis of the putative *Arabidopsis* DGAT amino acid sequence. (a) Comparison of acyltransferase sequences. The predicted amino acid sequence of the *Arabidopsis* DGAT (aDGAT) is shown in alignment with mouse DGAT (mDGAT) [9] and mouse ACAT (mACAT) [18]. Identical amino acid residues between the three proteins are highlighted and similar residues are shaded. (b) Hyrophobicity plot of the *Arabidopsis* DGAT amino acid sequence determined according to Kyte and Doolittle [19]. The hydrophobic domains lie above the dashed line.

into pFASTBAC for expression of the protein in Sf21 insect cell cultures. Lipids were extracted from cell cultures at 66 h after infection and resolved by TLC. Insect cells normally contain only small amounts of TAG [16] and it was clear that the cells infected with virus containing the putative DGAT encoding sequence had synthesized significantly more TAG than the cells infected with control virus (Fig. 2a). Cells infected with virus containing the DGAT coding sequence expressed a 51 kDa protein that was not present in cells infected with control virus (Fig. 2b). Biochemical activity of the DGAT was determined in vitro by measuring the incorporation of [^{14}C]DAG into TAG in the presence of oleoyl-CoA as acyl group donor. Insect cells expressing the DGAT cDNA had a greater than 8-fold higher DGAT activity than membranes from control cells (Fig. 2c). The synthesis of [^{14}C]TAG was strongly dependent on the addition of oleoyl-CoA since in its absence, the DGAT activity was almost 10-fold lower (Fig. 2c).

3.3. *DGAT gene structure, transcription start site and expression*

The *DGAT* gene corresponding to the cDNA described here

was recently sequenced on BAC F27F23 (http://genome-www.stanford.edu/Arabidopsis/). The gene is positioned on the upper arm of chromosome II close to the RI marker mi 139 at 35cM. The annotation of this genome sequence on the database suggested two open reading frames. However, the cDNA we sequenced shows that an intron/exon boundary was missed by the analysis program, that only a single gene is present and that it contains 15 introns. Southern blot analysis of *Arabidopsis* genomic DNA using the EST as a probe revealed the presence of only one *DGAT* gene in the *Arabidopsis* genome (data not shown). Primer extension analysis was conducted to determine the transcription start site of the *DGAT* gene (Fig. 3a). A single major reverse transcription product was observed co-migrated with a T residue at the −225 position relative to the putative translation start site (ATG). This transcription start site was confirmed using a second gene-specific oligonucleotide (PE2) for primer extension analysis. The gene contains a putative TATA box at nucleotide positions −353–−348 from the translation start site. The 5′ untranslated region (5′UTR) of the mRNA is 225 bases which is very much longer than the average distance for plant genes of 25–39 bases [17]. It remains to be seen whether the 5′UTR of the transcript from the *DGAT* gene is important in post-transcriptional regulation.

Northern blots revealed that the *DGAT* gene was expressed in *Arabidopsis* inflorescence and at a lower concentration in leaf (data not shown). In order to determine more easily the pattern of expression of the *DGAT* gene through a range of tissue types, RNA was isolated from the larger relative of *Arabidopsis*, *B. napus*. The mRNA was present at the highest concentrations in developing embryos, petals of flowers and developing flower buds (Fig. 3b) as expected since each of these tissues accumulates TAG. The *DGAT* mRNA was detectable but at much lower concentrations in leaf and stem, tissues which normally contain only 1–2% of the total cell lipid as TAG.

The cloning of a plant *DGAT* cDNA and identification of the gene will significantly help in experiments designed to determine the role and importance of DGAT in controlling the yield of storage oil in seeds. It may also aid our understanding of the role of TAG in other aspects of the plant life cycle such as pollen development.

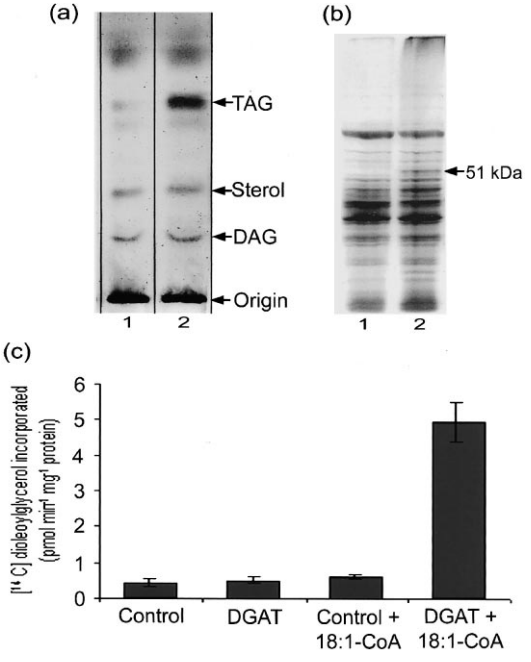

Fig. 2. Expression of the putative *DGAT* encoding sequence in insect cell cultures. (a) Neutral lipids were extracted from insect cells at 66 h after infection with wild-type baculovirus (1) and *Arabidopsis* DGAT recombinant baculovirus (2) and resolved by TLC. (b) Insect cells infected with wild-type baculovirus (1) and *Arabidopsis* DGAT recombinant baculovirus (2) were labelled with [^{35}S]methionine. The molecular masses were calculated from a standard curve using Broad Range markers (Bio-Rad). (c) [^{14}C]dioleoylglycerol incorporation into TAG was determined using membranes isolated from insect cells infected with wild-type baculovirus (1) and *Arabidopsis* DGAT recombinant baculovirus (2) ± oleoyl-CoA (18:1-CoA). Data represent the means of four separate experiments ($P < 0.001$)

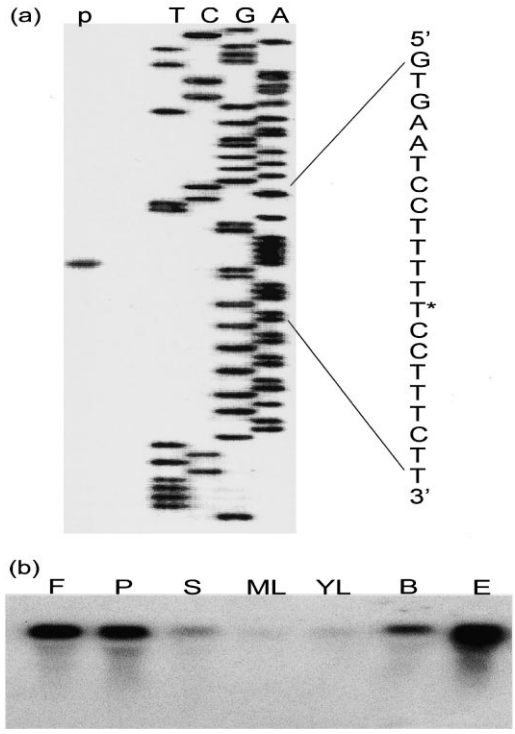

Fig. 3. Transcription start site of the *Arabidopsis DGAT* gene and mRNA concentrations of the *Brassica* homolog in rape tissues. (a) The product of the primer extension reaction was electrophoresed on a 6% sequencing gel (lane P) along with a *DGAT* 5′ flanking genomic DNA sequencing ladder (lanes T, C, G, A) generated by the same primer. The sense sequence around the cDNA product is shown on the right and the transcription start site is indicated by *. (b) RNA gel blots were hybridized with the *Arabidopsis DGAT* EST (E6B2T7). Lanes contain 20 μg RNA from open flower (F), petal (P), stem (S), mature leaf (ML), young leaf (YL), 6 mm bud (B), cotyledonary stage embryo (E).

Acknowledgements: We are grateful to Bob Farese and Sylvaine Cases of the Gladstone Institute, San Fransisco for a pre-print of their paper and Carole Thomas for her help with the insect cell cultures. We thank Denis Murphy and Steve Rawsthorne for their helpful comments on the manuscript. This work was supported financially by the Biotechnology and Biological Sciences Research Council (BBSRC) (UK), GAIT initiative.

References

[1] Focks, N. and Benning, C. (1998) Plant Physiol. 118, 91–101.
[2] Piffanelli, P., Ross, J.H.E. and Murphy, D.J. (1998) Sex. Plant Reprod. 11, 65–80.
[3] Wolters-Arts, M., Lush, W.M. and Mariani, C. (1998) Nature 392, 818–821.
[4] Ohlrogge, J. and Browse, J. (1995) Plant Cell 7, 957–970.
[5] Settlage, S.B., Wilson, R.F. and Kwanyuen, P. (1995) Plant Physiol. Biochem. 33, 399–407.
[6] Lacey, D.J. and Hills, M.J. (1996) Planta 199, 545–551.
[7] Perry, H.J. and Harwood, J.L. (1993) Phytochemistry 33, 329–333.
[8] Katavic, V. et al. (1995) Plant Physiol. 108, 399–409.
[9] Cases, S. et al. (1998) Proc. Natl. Acad. Sci. USA 95, 13018–13023.
[10] Verwoerd, T.C., Dekker, B.M. and Hoekema, A. (1989) Nucleic Acids Res. 17, 2362.
[11] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[12] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[13] Hills, M.J. and Mukherjee, K.D. (1990) Appl. Biochem. Biotechnol. 26, 1–10.
[14] Rost, B., Casadio, R., Fariselli, P. and Sander, C. (1995) Protein Sci. 4, 521–533.
[15] Ebinu, J.O., Bottorff, D.A., Chan, E.Y., Stang, S.L., Dunn, R.J. and Stone, J.C. (1998) Science 280, 1082–1086.
[16] Marheineke, K., Grunewald, S., Christie, W. and Reilander, H. (1998) FEBS Lett. 441, 49–52.
[17] Joshi, C.P. (1987) Nucleic Acids Res. 15, 6643–6653.
[18] Uelmen, P.J., Oka, K., Sullivan, M., Chang, C.C., Chang, T.Y. and Chan, L. (1995) J. Biol. Chem. 270, 26192–26201.
[19] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.