The Plant Cell, Vol. 7, 359–371, March 1995

Palmitoyl-Acyl Carrier Protein (ACP) Thioesterase and the Evolutionary Origin of Plant Acyl-ACP Thioesterases

Aubrey Jones, H. Maelor Davies, and Toni A. Voelker[1]

Calgene, Inc., 1920 Fifth Street, Davis, California 95616

Acyl-acyl carrier protein (ACP) thioesterases play an essential role in chain termination during de novo fatty acid synthesis and in the channeling of carbon flux between the two lipid biosynthesis pathways in plants. We have discovered that there are two distinct but related thioesterase gene classes in higher plants, termed *FatA* and *FatB*, whose evolutionary divergence appears to be ancient. *FatA* encodes the already described 18:1-ACP thioesterase. In contrast, *FatB* representatives encode thioesterases preferring acyl-ACPs having saturated acyl groups. We unexpectedly obtained a 16:0-ACP thioesterase cDNA from *Cuphea hookeriana* seed, which accumulate predominantly 8:0 and 10:0. The 16:0 thioesterase transcripts were found in non-seed tissues, and expression in transgenic *Brassica napus* led to the production of a 16:0-rich oil. We present evidence that this type of *FatB* gene is ancient and ubiquitous in plants and that specialized plant medium-chain thioesterases have evolved independently from such enzymes several times during angiosperm evolution. Also, the ubiquitous 18:1-ACP thioesterase appears to be a derivative of a 16:0 thioesterase.

INTRODUCTION

In plants, de novo fatty acid synthesis is located exclusively in the stroma of plastids, where the acyl chains are covalently bound to a soluble acyl carrier protein (ACP) during the extension cycles. Carbon chain elongation can be terminated by transferring the acyl group to glycerol 3-phosphate, thereby retaining it in the plastidial, "prokaryotic," lipid biosynthesis pathway. Alternatively, specific thioesterases can intercept the prokaryotic pathway by hydrolyzing the newly formed acyl-ACP into free fatty acids and ACP. Subsequently, the free fatty acids exit the plastids by an undetermined mechanism and supply the "eukaryotic" lipid biosynthesis pathway. The latter is located in the endoplasmic reticulum and is responsible for the formation of phospholipids, triglycerides, and other neutral lipids. Therefore, by catalyzing the first committed step in the eukaryotic lipid biosynthesis pathway in all plant cells, acyl-ACP thioesterases play a crucial role in the distribution of de novo synthesized acyl groups between the two pathways (Löhden and Frentzen, 1988; reviewed in Browse and Somerville, 1991; Gibson et al., 1994).

To date, commonly found plant acyl-ACP thioesterases are specific for 18:1-ACP and therefore have been termed "long-chain" or "oleoyl"-ACP thioesterases. Encoded by nuclear genes, 18:1-ACP thioesterases are synthesized as preproteins in the cytoplasm and subsequently imported into the plastids. Sequences have been obtained from several angiosperm families, and aside from high variability in the transit peptides, they are very similar to each other (Hitz and Yadav, 1992; Knutzon et al., 1992; Dörmann et al., 1994, 1995). The 18:1-ACP thioesterase is considered an essential "housekeeping" enzyme in all plant cells because 18:1 together with its derivatives usually represents the bulk of de novo fatty acyl carbon produced (Gibson et al., 1994).

We previously isolated a 12:0-ACP thioesterase and obtained a cDNA from 12:0-producing seed of California bay trees (*Umbellularia californica*, Lauraceae; Davies et al., 1991; Voelker et al., 1992). When expressed in developing oil seed of Arabidopsis, the bay enzyme redirects the resident long-chain fatty acid synthase to 12:0 production. This demonstrated for the first time the chain length–determining role of acyl-ACP thioesterases during de novo fatty acid biosynthesis (Voelker et al., 1992). The 12:0-ACP thioesterase was found to share significant sequence homology with 18:1-ACP thioesterases of other species (Knutzon et al., 1992); this suggests that during the evolution of California bay trees, a housekeeping 18:1-ACP thioesterase gene was recruited for the development of an enzyme with shorter chain length specificity.

Seed reserve lipids containing medium-chain (C8 to C14) fatty acids are not restricted to the Lauraceae but are also found in a limited number of taxonomically dispersed families of angiosperms (Hilditch and Williams, 1964). Biochemical evidence for medium-chain acyl-ACP thioesterases was found in medium-chain–producing seed of plant families other than the Lauraceae (Davies, 1993; Dörmann et al., 1993). Therefore, it seemed likely that functional homologs of the bay thioesterase, with specificities corresponding to the fatty acid profile of the respective seed oils, are present in such species. Localized sequence conservation between 18:1- and the 12:0-ACP

[1] To whom correspondence should be addressed.

thioesterases facilitated a general evolutionary polymerase chain reaction (PCR) strategy for the cloning of cDNAs encoding such enzymes.

The genus *Cuphea* (Lythraceae) is unusual because its members contain seed lipids whose diverse compositions emphasizing medium-chain lengths are unmatched elsewhere in the plant kingdom (Graham et al., 1981; Graham and Kleiman, 1987). One species, the Mexican shrub *C. hookeriana*, accumulates predominantly 8:0 and 10:0 in its seed lipids (Graham et al., 1981). In applying the evolutionary PCR-based cDNA cloning strategy to this species, we discovered an acyl-ACP thioesterase with unexpected substrate specificity. Here, we describe the cloning of a cDNA and the characterization of its encoded 16:0-ACP–preferring thioesterase, and we demonstrate its impact on the fatty acid synthases of *Escherichia coli* and plants. In addition, by analyzing sequence data, we discovered that plant thioesterases fall into two evolutionary lineages: 18:1 thioesterases and a newly defined second clade, which contains all known plant medium-chain thioesterases and the newly discovered 16:0 thioesterase, which we consider ubiquitous in plants. We discuss the impact of these findings on our current understanding of plant lipid biosynthesis.

RESULTS

Generation of *C. hookeriana* cDNA Fragments Based on Sequences Conserved between Safflower and Bay Thioesterases

Following the isolation of a cDNA encoding an enzyme involved in 12:0 production in California bay (Voelker et al., 1992), we were interested in generating cDNAs for homologous enzymes in medium-chain–producing species of other plant families. For one RNA source, we used maturing seed of *C. hookeriana*, a member of the Lythraceae whose seed reserve triglycerides contain 60 mol% 8:0 and 25% 10:0 (Graham et al., 1981). We applied a two-stage cloning protocol. First, single-stranded cDNA was synthesized, and thioesterase sequences were amplified from this template using a mixed oligonucleotide PCR strategy (Gould et al., 1989). Primers were designed using six different five- to seven-residue amino acid sequences conserved between the safflower 18:1-ACP thioesterases *Ct* FatA1 and *Ct* FatA2 ("Fat" denotes plant fatty acyl-ACP thioesterase; see Methods and Table 1 for terminology and sources) and the California bay 12:0-ACP thioesterase *Uc* FatB1 (see alignment in Figure 1). Two of the primer sets amplified DNA of the expected length, namely, the reactions using S and V (~300 bp) and V and X (~110 bp) primer sets as anchors (Figure 1, bars atop the alignment).

Two similar but distinct sequences from the S and V reaction and one from the V and X reaction were obtained after the cloning of amplified DNA. As shown in Figure 1, their deduced amino acid sequences (SV1, SV2, and VX1) are all related to the two reference thioesterases. Both S and V sequences are significantly closer to *Uc* FatB1 (~65% identity) than to *Ct* FatA2 (only ~40% identity). Also, they share a four–amino acid deletion (Δ1 in Figure 1) with *Uc* FatB1 relative to *Ct* FatA2. The VX1 sequence is equally distant from both reference sequences (50% identity).

Table 1. Sources of Thioesterase Sequences

Species	Mnemonic	Previous Names/ Sources	GenBank No.
Arabidopsis thaliana	*At FatA1*	TE 1-7[a]	Z36912
	At FatB1	TE 3-2[a]	Z36911
Brassica napus	*Bn FatA1*	NL2[b]	X73849
Brassica rapa	*Br FatA1*	NA	U17098
Cuphea hookeriana	*Ch FatA1*	NA[c]	
	Ch FatB1	NA[d]	U17076
	Ch FatB2	NA[c]	
Cuphea lanceolata	*Cl FatB1*	*Cl*/TE13[e]	X76561
Coriandrum sativum	*Cs FatA1*	cors[f]	L20978
Carthamus tinctorius	*Ct FatA1*	2-1[g]	M96569
	Ct FatA2	5-2[g]	M96568
Glycine max	*Gm FatA1*	2[h]	
	Gm FatA2	4[h]	
Ulmus americana	*Ua FatB1*	NA[i]	
Umbellularia californica	*Uc FatB1*	BTE[j]	M94159
	Uc FatB2	NA[k]	U17097

For rules of nomenclature, see Methods.
[a] Dörmann et al. (1995).
[b] Loader et al. (1993).
[c] K. Dehesh, D. Knutzon, A. Jones, and T.A. Voelker (unpublished data).
[d] This study.
[e] Töpfer and Martini (1994).
[f] Dörmann et al. (1994).
[g] Knutzon et al. (1992).
[h] Hitz and Yadav (1992).
[i] D. Knutzon, A. Jones, and T.A. Voelker (unpublished data).
[j] Voelker et al. (1992).
[k] A. Cranmer and T.A. Voelker (unpublished data).
NA, no name available.

Cloning of a *C. hookeriana* Thioesterase cDNA

In the second cloning stage, the SV1 PCR product was used as a probe for the screening of *C. hookeriana* maturing seed cDNA libraries. Of 250,000 primary plaques, we isolated five *Ch FatB1* clones. The longest cDNA, the 1744-bp pCGN4806, will be described here in detail (DNA sequence not shown; GenBank accession number U17076). Its open reading frame extends from base pair 47 to base pair 1450, and the first in-frame translational start codon is located at base pair 206. Surrounding this ATG is a sequence resembling the plant consensus sequence for initiation of translation (Lütcke et al., 1987), and a 416–amino acid polypeptide can be derived (Figure 1, *Ch* FatB1). Aside from mutations in the S primer coding region, the SV1 sequence is contained in *Ch* FatB1 and therefore was

probably derived from the same gene. Between base pairs 370 and 1300, the *Ch FatB1* cDNA is essentially colinear with the California bay cDNA *Uc FatB1*. In this region, the two DNA sequences average 61% identity. Sequence identities between *Ch FatB1* and safflower *Ct FatA2* are generally lower and more restricted, spanning base pairs 500 to 1200 of *Ch FatB1* (data not shown).

Aligning the derived *Ch* FatB1 amino acid sequence with those of *Uc* FatB1 and *Ct* FatA2 showed the trend already observed with DNA sequences—*Ch* FatB1 is more similar to *Uc* FatB1 than to *Ct* FatA2 (Figure 1). To clarify further, we also present these alignments graphically (Figure 2A). Significant sequence identities with *Uc* FatB1 begin approximately at amino acid 90 of *Ch* FatB1, a position homologous to amino acid 60 of *Uc* FatB1. This precedes the N terminus of the purified *Uc* FatB1 protein (Voelker et al., 1992) by 24 amino acids (Figure 1, amino acid 84; underlined). Significant sequence identities continue to ~20 amino acids from the C termini. In summary, aside from length variations in the transit peptides and sequence deviation at the very C termini, both polypeptides are essentially colinear. Less sequence identity is found when *Ch* FatB1 and *Ct* FatA2 are aligned (Figure 2A, thin lines). The first significant identities begin only after amino acid 140 of *Ch* FatB1, and the frequency and extent of highly conserved areas are reduced. Also, as indicated by a shift in the plot, *Ch* FatB1 has ~20 amino acids fewer than *Ct* FatA2 in the C-terminal part of the molecule (Δ2, Figure 1). Overall, *Ch* FatB1 is much more similar to *Uc* FatB1 than to *Ct* FatA2.

Ch FatB1 Encodes a 16:0-ACP Thioesterase

We previously found that *E. coli*–expressed "mature" *Uc* FatB1 was very active in vitro and had an acyl-ACP chain length specificity identical to that of the authentic plant enzyme (Voelker and Davies, 1994). Accordingly, *Ch* FatB1 was expressed in *E. coli* in an analogous manner, and crude extracts of cultures were assayed for acyl-ACP hydrolytic activity in vitro (Figure 3). The extract of the *Ch* FatB1–expressing strain showed a much greater preference for 16:0-ACP as substrate than did the control lysate. There were also slight but reproducible increases in the activities of 14:0-, 18:0-, and 18:1-ACP. No difference between activity profiles of control and *Ch* FatB1 was found when acyl-CoA substrates were used in the assay (data not shown). These results suggested that *Ch* FatB1 could be an acyl-ACP thioesterase active on C14 to C18 chains, with strong preference for 16:0-ACP.

The enzymatic specificity of *Ch* FatB1 was completely unexpected because seeds of *C. hookeriana* do not accumulate significant amounts of fatty acyl groups of chain lengths greater than C10. Therefore, *Ch* FatB1 appeared not to be involved in the production of 8:0 and 10:0 for storage lipids; this conclusion

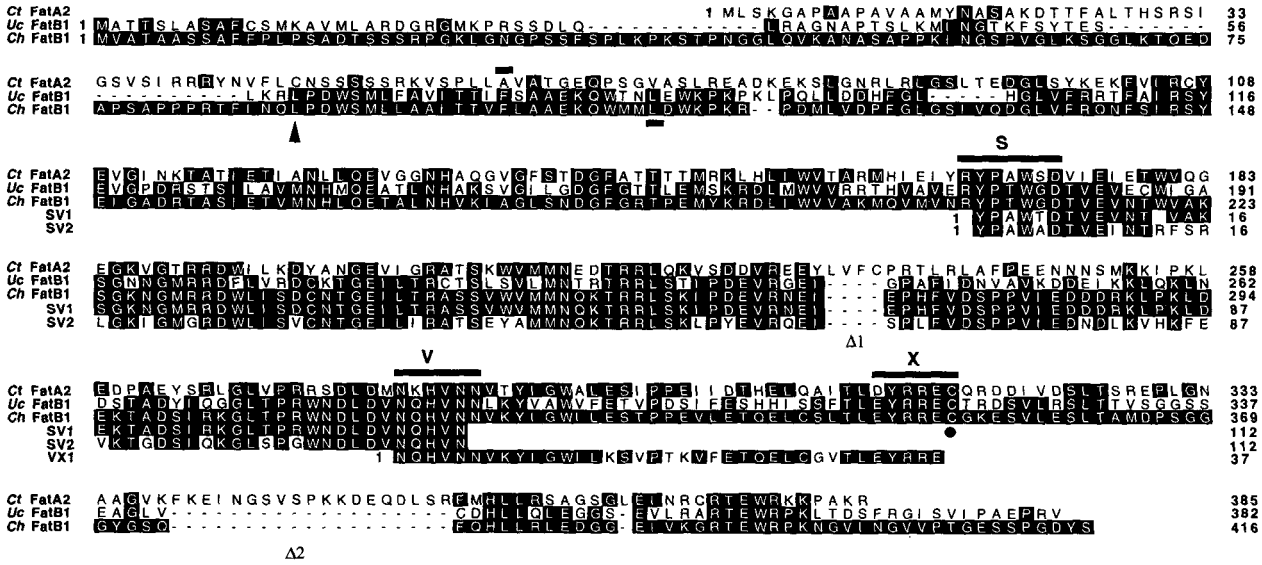

Figure 1. Alignment of Thioesterase Amino Acid Sequences.

The derived amino acid sequence of the predicted thioesterase preprotein *Ch* FatB1 is aligned with an 18:1 thioesterase of safflower, *Ct* FatA2 (Knutzon et al., 1992), and the 12:0 thioesterase of California bay, *Uc* FatB1 (Voelker et al., 1992). To improve matches, gaps were introduced in the respective sequences. All residues identical to *Ch* FatB1 are shown in black. Bars indicate the N-terminal amino acid residues of *Ct* FatA2 (amino acid 61) and *Uc* FatB1 (amino acid 84), as determined by sequencing purified proteins. The newly proposed mature N terminus of FatB thioesterases is indicated by an arrowhead. Symbols Δ1 and Δ2 indicate sequence variations indicative of the Fat subclasses. The bars labeled S, V, and X indicate the sequences used successfully for PCR. The derived amino acid sequences of three *C. hookeriana* PCR products, SV1, SV2, and VX1, are included. The proposed cysteine active site is marked (●).

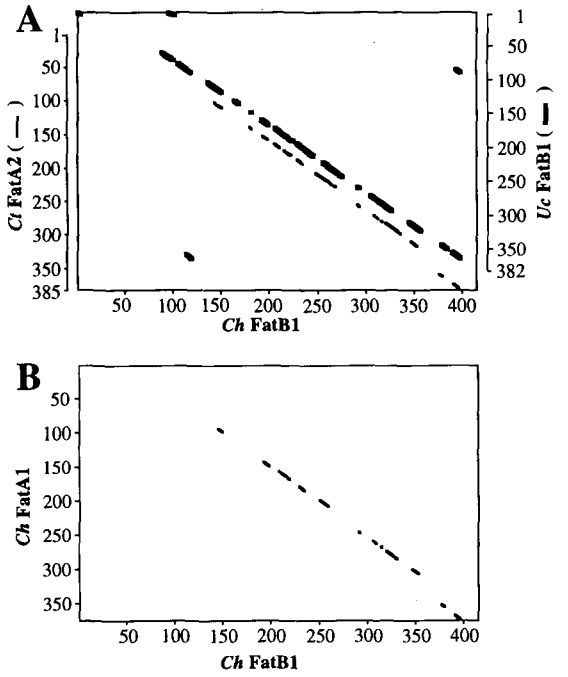

Figure 2. Dot Matrix Alignments of Thioesterase Polypeptides.

For all alignments, we used the same stringency (window of eight amino acids; minimum identity needed to score was 80%).
(A) Alignment of *Ch* FatB1 with reference thioesterases. *Ch* FatB1 was aligned with the 18:1-ACP thioesterase *Ct* FatA2 and subsequently with the 12:0-ACP thioesterase *Uc* FatB1. The results of both alignments are shown in one graph. To avoid overlap of signals, the coordinates of the two reference sequences were staggered relative to each other. Results obtained with the FatA reference are shown by thin lines, and results with the FatB reference are shown by thick lines.
(B) Alignment of two *C. hookeriana* isolates, *Ch* FatA1 and *Ch* FatB1.

was supported by our finding of *Ch FatB1* transcripts in flowers, leaves, and roots of *C. hookeriana* at levels comparable to those in the maturing seed (data not shown).

Sequence Comparisons Define Two Classes of Plant Thioesterases

There is a greater difference between the *C. hookeriana Ch* FatB1 and safflower *Ct* FatA2 sequences than there is between *Ch* FatB1 and California bay *Uc* FatB1 sequences (Figures 1 and 2A). This simply could reflect the phylogenic relationships of these three species. However, a consideration of additional Fat sequences showed that this interpretation is too simple. K. Dehesh, A. Jones, and T. A. Voelker (unpublished data) obtained another *Fat* cDNA from *C. hookeriana*. Although one might expect its derived polypeptide, *Ch* FatA1, to be very similar to *Ch* FatB1, it is in fact quite distant (30% conservation; see alignment in Figure 2B). The combination of Figures 2A and 2B shows that *Ch* FatB1 is approximately equally distant from the FatA sequence of its own species and from *Ct* FatA2 of safflower. In contrast, *Ch* FatA1 is closely similar to *Ct* FatA2 (64%; data not shown). This suggests a different ancestry for the *FatA* and *FatB* genes of *C. hookeriana*. It appears that Fat sequences are grouped into FatA or FatB independent of the taxonomic relationships of the plants from which they come.

Based on these sequence interrelations, an evolutionary scenario can be suggested. It is evident that the most recent time of divergence of the lineages leading to *Ch* FatB1 and California bay *Uc* FatB1 would have been the date of divergence of the respective families (Lythraceae and Lauraceae). By the same token, the most recent time of separation of the lineages leading to *C. hookeriana Ch* FatA1 and safflower *Ct* FatA2 would have been the date of divergence of Lythraceae and Asteraceae. Because FatA and FatB members of *C. hookeriana* are much less related to each other than are interspecies members within each sequence group, the divergence leading to FatA and FatB must have substantially predated the family divergences. Therefore, the last common ancestor of these three families must have already carried two distinct thioesterase sequence classes. It can also be inferred that all extant descendants of this common ancestor should still carry the two sequence types of thioesterases (except when lost secondarily). Multiple support for this scenario appeared (see following discussion).

In the course of systematic sequencing of an Arabidopsis developing-silique cDNA library, a 463-bp DNA fragment, which

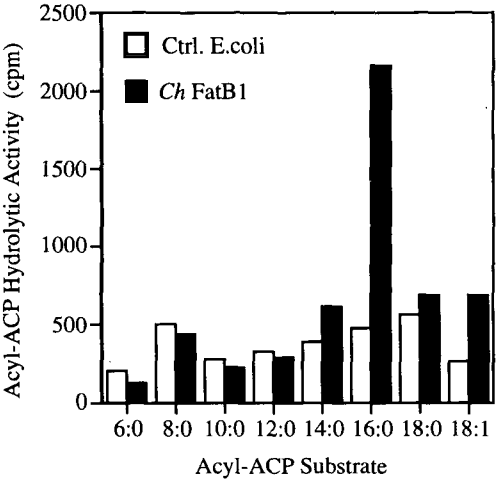

Figure 3. *Ch* FatB1 Is a 16:0-ACP–Preferring Thioesterase.

E. coli was transformed with vector only (Ctrl.) or the *Ch* FatB1 expression plasmid. For determination of the in vitro acyl-ACP hydrolytic activities, cells of isopropyl β-D-thiogalactopyranoside–induced logarithmically growing cultures were sonicated. After removal of debris by centrifugation and adjustment to equal total protein concentration, equivalent aliquots of the crude extracts were assayed. The graph shows the mean values from three replicates. Variations did not exceed ±2.5%.

appeared to code for a *Uc* FatB1 homolog, was obtained (Grellet et al., 1993). This fortuitous cloning validated our hypothesis that *FatB* genes may also exist in species that do not accumulate medium-chain fatty acids. The DNA fragment of Grellet et al. (1993) allowed Dörmann et al. (1995) to clone a corresponding full-length cDNA from Arabidopsis. Its derived amino acid sequence is close to that of *Ch* FatB1 (65% conservation; data not shown) and therefore falls into the FatB class (here named *At* FatB1).

To obtain additional support for our two-thioesterase hypothesis, we compared a total of 16 thioesterase cDNAs from 10 species of seven different angiosperm families. Because DNA sequence contains more phylogenic information than its derived amino acid sequence (Chase et al., 1993), we chose to base our cladistic analysis on homologous DNA regions shared by all members of the sample. This restricted the comparison to an ~1000-bp region coding for most of the so-called mature portion of the derived polypeptides (residues 90 to 400 for *Ch* FatB1; Figure 2). In this 1000-bp region, the DNA sequences of the sample are practically colinear, and scores of 40 to 85% sequence identity were obtained after the introduction of a moderate number of gaps for better alignment (data not shown). A quantitative search yielded a single-most parsimonious tree (Figure 4).

A branching at the origin of the tree separates *FatA* from *FatB* sequences (Figure 4). When we based the search on various 300- to 600-bp DNA fragments contained within the 1000-bp region, this principal branching was always replicated, and no member ever switched from one class to the other. This was evidently the first evolutionary division of the genes of the sample. The *Ct FatA2* and *Ch FatA1* sequences reside together with 18:1-ACP thioesterase cDNA sequences from soybean (*Gm* denotes *Glycine max*; Fabaceae) and coriander (*Cs* denotes *Coriandrum sativum*; Umbelliferae) in the "upper" clade (A). Two other *C. hookeriana* cDNAs, including *Ch FatB1*, reside in the "lower" clade (B). A cDNA isolated from *Cuphea lanceolata* (*Cl FatB1*; Töpfer and Martini, 1994) is closest to *Ch FatB1*. Also, *Uc FatB1* and the *FatB* representatives from Arabidopsis (*At FatB1*) and elm (*Ua* deonotes *Ulmus americana*; Ulmaceae) reside in the B clade.

This DNA-based analysis of sequence relatedness supports our notion of two thioesterase classes, which we originally deduced from amino acid sequences. Note that representatives of both sequence classes were isolated from Arabidopsis and from *C. hookeriana*, even though only the latter species is known to synthesize significant amounts of medium-chain fatty acids. These two pairs represent the most significant support for our hypothesis that the basic division into *FatA* and *FatB* sequence lineages reflects neither taxonomic relationships nor seed acyl compositions. Thus, *FatA* and *FatB* genes are paralogous; that is, they are derived from an ancient gene duplication (Fitch, 1970). In turn, the creation and subsequent conservation of two distinct thioesterase sequence classes in many lineages suggest that the encoded enzymes may have different functions.

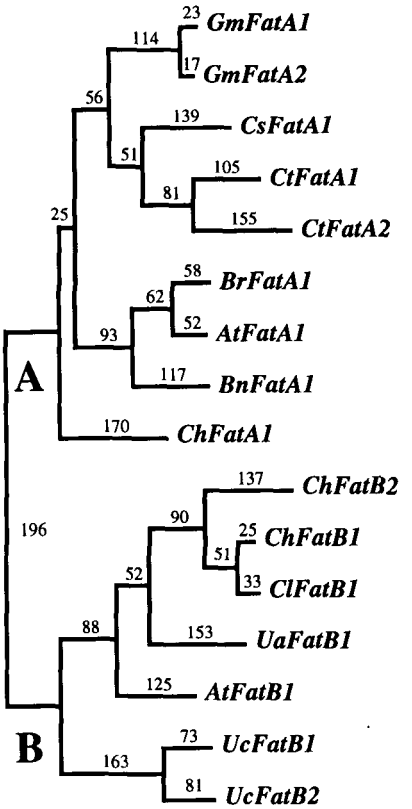

Figure 4. Phylogenic Tree Based on Plant Thioesterase DNA Sequences.

To minimize random errors, we based our search on a *Fat* region without extensive length variations or questionable homology. This restricted us to an ~1000-bp region covering the majority of the mature polypeptide coding frame. The resulting multiple DNA sequence alignment was used to search for the most parsimonious trees. The quantitative branch-and-bound version of the computer program PAUP 3.0 (Swofford, 1991) was used. Numbers at the branches reflect the nucleotide divergence of the respective branch from the most recent common ancestor sequence. Sources of the respective DNA sequences are listed in Table 1. A and B denote the FatA and FatB subtrees.

Ch FatB1 Redirects Bacterial and Plant Fatty Acid Synthases to Palmitate and Myristate Production

With respect to structure, the fatty acid synthases of plants and the eubacterium *E. coli* are classified as prokaryotic or type II. Each individual fatty acid synthesis reaction is catalyzed by a discrete, monofunctional enzyme. These structural, and also sequence, similarities of the enzymes involved (Magnuson et al., 1993) are vestiges of a shared ancestry, with the chloroplast essentially a descendant of eubacterial cyanobacteria (Gray, 1989). We therefore tested whether plant thioesterases can function in a bacterial fatty acid synthase. By expression of "mature" *Uc* FatB1 in *E. coli*, which caused

the accumulation of large quantities of free medium-chain fatty acids in the growing cultures, Voelker and Davies (1994) demonstrated the compatibility.

To characterize *Ch* FatB1 further, we analyzed its in vivo action in the bacterial system. Figure 5A shows that the control culture accumulated only traces of free fatty acids. More than 95% of its total acyl groups (60 nmol/mL; data not shown) were combined in phospholipids. The introduction of *Ch* FatB1 increased total fatty acid accumulation drastically, to 155 nmol/mL (same cell density), and the increase was found in free fatty acids, as shown in Figure 5A. The most prominent free fatty acids were 14:0 and 16:0, but also some 16:1 and 18:1 were produced. These free fatty acids accumulated most likely as a direct result of *Ch* FatB1's hydrolysis of acyl-ACP intermediates. Two peculiarities of the *E. coli* fatty acid synthase (reviewed in Magnuson et al., 1993) must be considered when interpreting this result: 16:0-ACP is normally not extended, so the potential in vivo 18:0-ACP hydrolytic activity of *Ch* FatB1 cannot be tested in this system. Also, unsaturated acyl-ACPs with chain lengths C10 to C16 are present, in contrast to plastids. The considerable 14:0 production in vivo in relation to the enzyme specificity (Figure 3) can be explained if there was sufficient enzyme present in vivo to consume most of the carbon flux at the 14:0-ACP stage, thereby depleting the 16:0-ACP substrate pool available for hydrolysis. Indeed, we found that 50% of the flux through 14:0-ACP was diverted by hydrolysis, as calculated from accumulation data. This apparent mismatch between in vitro enzyme specificity and in vivo fatty acyl production was investigated previously (Davies, 1993; Voelker and Davies, 1994). In summary, these results support the in vitro characterization of *Ch* FatB1 as a thioesterase acting on acyl-ACPs with chain lengths C14 to C18, with preference for saturated substrates.

We previously showed that *Uc* FatB1 results in 12:0 production in the seed of the crucifer Arabidopsis (Voelker et al., 1992). Here, we report the expression of the 16:0-preferring *Ch FatB1* in a canola variety of the cruciferous oil crop *Brassica napus.* The reading frame of *Ch FatB1* was inserted into a maturing seed-specific expression cassette derived from a *B. rapa* napin gene (Kridl et al., 1991). Subsequently, *B. napus* was transformed with the chimeric gene by Agrobacterium cocultivation (Radke et al., 1988), and plants were regenerated. Seed development of the *Ch FatB1* transformants appeared unaffected, but the 16:0-ACP hydrolytic activities in extracts of the maturing seed were elevated up to threefold relative to that of control plants (data not shown).

We also determined the lipid composition of mature seed from transgenic plants. The canola variety used for transformation accumulated predominantly C18 fatty acids (Figure 5B), with only traces of 12:0 (0.02 mol%), 14:0 (0.14 mol%), some 16:0 (6 mol%), and 16:1 (0.5 mol%). Minor amounts of very long chains (C20 or longer) were also present. In the seed of 25 independent *Ch FatB1* transformants, 16:0 ranged from 7 to 34 mol% (single example shown in Figure 5B). Palmitate levels in pooled seed of individual transformants were distributed predominantly between 13 and 27 mol%. Quantitative

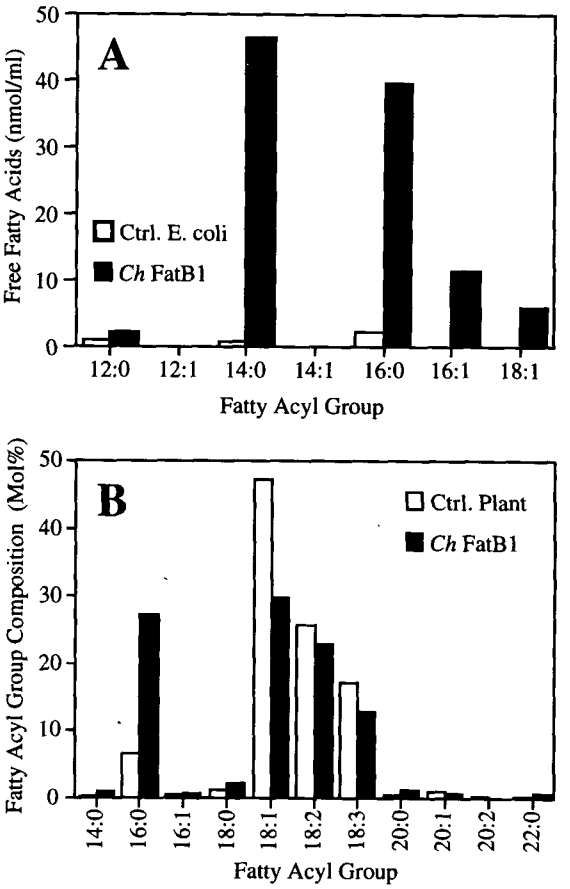

Figure 5. Redirecting Fatty Acid Production in *E. coli* and Canola Seed.

(A) Production of free fatty acids in *E. coli.* The fatty acid degradation mutant strain K27 (*fadD*) was transformed with vector only (Ctrl.) or the *Ch* FatB1 expression plasmid. Samples were taken after the cultures had reached 5 × 10^8 cells per milliliter.

(B) Impact of *Ch FatB1* expression in maturing canola seed on the fatty acid composition of seeds. Twenty mature seeds each of a canola control plant and a transformant (*Ch FatB1* plant 13) were pooled for fatty acyl determination.

plant-to-plant variation of the induced phenotype, caused by transgene copy number and genomic location, is to be expected (Voelker et al., 1992). The quantity of 14:0 was multiplied (from 0.14 to 1 mol%), but also 18:0 and 20:0 increased, and these increases correlated with the 16:0 levels (data not shown). However, there was no measurable production of 8:0 or 10:0, confirming our conclusion that *Ch* FatB1 is probably not the enzyme responsible for formation of these fatty acids in *C. hookeriana* embryos. The responsible thioesterase is *Ch* FatB2 (K. Dehesh, A. Jones, D. Knutzon, and T.A. Voelker, unpublished data; see also Figure 4). The absolute increase in 16:0 content of the seed (data not shown) was balanced predominantly by a reduction in 18:1 content, a situation similar to that observed previously when 12:0 was produced by *Uc* FatB1 in Arabidopsis (Voelker et al., 1992).

DISCUSSION

Ch FatB1 Enables the Genetic Engineering of High-Palmitate Oils

Palmitate usually represents only 5 to 10% of the fatty acid composition of most seed oils, although it can be as high as 77%. Palm mesocarp is currently the predominant bioresource for 16:0-rich (32 to 47%) oils (Hilditch and Williams, 1964). There is much interest in increasing the 16:0 content of the predominantly C18-producing temperate oilseed crops; this would improve texture in margarines and shortenings derived from these sources (Safford et al., 1993). Although conventional breeding and mutagenesis schemes have been employed, there has been only moderate success (Persson, 1985; Bubeck et al., 1989). When *Ch FatB1* was expressed in the seed of transgenic canola, the 16:0 levels increased from 6% up to 35% of total. During maturation of these seeds, the in vitro 16:0-ACP hydrolytic activity was increased, most likely caused by *Ch* FatB1 accumulation. The overall hydrolytic profile was reminiscent of maturing palm mesocarp (Sambanthamurthi and Oo, 1990). It is tempting to speculate that a *Ch* FatB1–like enzyme is overexpressed naturally in this tissue.

It is interesting that *Ch FatB1* expression in *E. coli* and maturing plant seed induced fatty acid phenotypes that are comparable in quantitative terms to *Uc FatB1* expression under comparable conditions, yet in vitro–measured *Ch* FatB1 acyl-ACP hydrolytic activities in both organisms were hundreds of times lower than that found with *Uc FatB1* (Voelker et al., 1992; Voelker and Davies, 1994). We cannot explain this unexpected discrepancy. In summary, our high-palmitate canola illustrates the continuing utility of plant acyl-ACP thioesterases for the useful modification of oilseed crops.

Two Sequence Classes of Plant Acyl-ACP Thioesterases

Our comparison of 16 plant acyl-ACP thioesterase cDNA sequences (Figure 4) shows that they fall into two distinct classes, which we have termed *FatA* and *FatB*. Nine *FatA* cDNA sequences, isolated from five different plant families (Table 1), encode 363 to 390 residue preproteins. The length variation, almost exclusively restricted to the N-terminal region, is normal for stromal transit peptides, which these portions most likely represent (de Boer and Weisbeek, 1991). In contrast, the mature portions of the polypeptides are highly conserved in length and sequence; a region homologous to residues 89 to 327 in *Ct* FatA2 (Figure 1) is conserved at least 80% between all FatA representatives.

Seven *FatB* cDNAs were isolated from four different plant families (Table 1). Preproteins range from 382 to 414 residues, and excepting representatives isolated from a single species or genus, such as *U. californica* or *Cuphea*, the sequences of an ~60– to 90–amino acid N-terminal portion are not conserved (data not shown). Significant similarity starts abruptly with an almost completely conserved region (*Uc* FatB1 residues 60 to 82; Figure 1), and conservation averaging 70 to 80% continues almost to the C terminus. All FatB sequences share the Δ1 and Δ2 length variation relative to FatA (Figure 1); this indicates that these features evolved very soon after the *FatA*/*FatB* divergence and can be considered markers for class specificity. Conserved deletion/insertion events have been implicated as reliable indicators of phylogenic relationships (e.g., Morden and Golden, 1989).

The advent of more FatB representatives provided an explanation for a previously puzzling feature of *Uc* FatB1. The mature N terminus of the purified *Uc* FatB1 protein was originally determined by amino acid sequencing as residue 84 (underlined in Figure 1; Voelker et al., 1992). However, this would place the most N-terminal region conserved among all FatB representatives (residues 60 to 82 for *Uc* FatB1; Figure 1) into the transit peptide. This sequence would be unusual for stromal transit peptides not only because it is highly conserved, but also because it contains a hydrophobic domain resembling thylakoid transit peptides (de Boer and Weisbeek, 1991). We now believe that this conserved region is not part of the transit sequence and that it is included in the N-terminal portion of the processed mature protein. Consistent with this hypothesis, expression of *Uc FatB1* in transgenic canola seed produced a *Uc* FatB1 antigen having an M_r of 40 kD, that is, larger than the purified seed protein (34 kD) (SDS-PAGE; Voelker et al., 1992; T.A. Voelker, unpublished data). We believe that the 40-kD form represents the in vivo state of *Uc* FatB1 and that limited N-terminal proteolysis occurred during purification. This reduces the putative transit peptide to ≤60 residues. Also, a 40-kD form for *At* FatB1 has been detected in Arabidopsis tissues (Dörmann et al., 1995). Most other FatB sequences have 80 to 90 amino acids upstream of the hydrophobic domain. Either these propeptides have longer transit sequences or are cleaved even more closely to the N terminus.

Interestingly, the hydrophobic domain is absent from FatA sequences. Because it appears nonessential for enzyme activity in vitro and in vivo (Figures 3 and 5A; T.A. Voelker, unpublished data), its role remains unexplained. Perhaps FatB is loosely anchored to a membrane or complexed with the other enzymes of the fatty acid synthase via this domain. Only one cysteine (close to the C terminus, indicated by a dot in Figure 1) is conserved between all Fat sequences to date, making it a likely candidate for the active site residue. Inhibitor studies by Pollard et al. (1991) and Knutzon et al. (1992) showed that FatA and FatB members are sensitive to cysteine inactivation.

The *FatA*/*B* Divergence Predates Angiosperm Radiation

From the sequence-based tree, we concluded that the *FatA* and *FatB* lineages must have separated before the divergence of any of the plant families under consideration. To determine the age of this separation, we needed information on the

evolutionary relationships of all these plant families. The most extensive attempt at mapping a phylogenic tree of seed plants is that of Chase et al. (1993). Their study was based on DNA sequences coding for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, and they considered representatives of more than 150 seed plant families. This phylogenic analysis suggests that the Lauraceae lineage appears to have separated very early in angiosperm radiation from the lineage leading to all other plant families (Chase et al., 1993) used in our study. This pattern is reproduced in the *FatB* subtree of Figure 4; this pattern of relationships separates the *U. californica* sequences from all other *FatB* representatives. Because the separation of *FatA* and *FatB* must predate the split of the Lauraceae lineage from all others, both *Fat* classes must have existed early in the radiation of the angiosperms, which is dated to 150 to 300 million years ago (see Martin et al., 1993; reviewed in Sanderson and Donoghue, 1994).

The positioning of a given *Fat* into the *FatA* and *FatB* subclasses, as obtained by parsimony analysis, was reconfirmed using alternative criteria. For example, a tree based on sequence similarity, but not parsimony, yielded the same basic division as that shown in Figure 4 (using the Clustal software program [Higgins and Sharp, 1989], which makes no a priori assumption of relatedness, but groups sequences by examining the distances between all pairs; tree not shown). Also, several large deletions and insertions define the two subclasses, as described earlier in this study. In summary, *FatA* and *FatB* are quite diverged, significantly more so than the divergence observed between the California bay *FatB* sequences and any other *FatB* representative. Taken together, there is clear evidence that (1) the rooting of the evolutionary tree is solid, and (2) the *FatA/B* gene duplication is much older than the Lauraceae divergence, as discussed previously. This makes it likely that both thioesterase classes are also present in the gymnosperms and possibly also in more distant taxa of plants (Sanderson and Donoghue, 1994). Furthermore, *FatA* and *FatB* genes should be present in all species of these taxa unless secondarily lost.

The Two Thioesterase Classes Are Also Functionally Distinct

The acyl-ACP hydrolytic specificities of five FatA representatives from three families have been measured in vitro after expression in *E. coli.* All appeared to be 18:1 specific, with minor activities on 18:0 and 16:0 substrates (Hitz and Yadav, 1992; Knutzon et al., 1992; Dörmann et al., 1994). This suggests that FatA is involved predominantly in the hydrolysis of 18:1-ACP. Because all five isolates from three different plant families gave a near-identical hydrolytic specificity, it is likely that FatA plays this conserved 18:1-ACP thioesterase role in all higher plants. This is a critical function in plant fatty acid biosynthesis as 18:1 is the immediate precursor of most fatty acids found in phospholipids and triglycerides synthesized in the eukaryotic pathway (Gibson et al., 1994).

Contrasting with the FatA conservation, the specificities of FatB enzymes show high variability. The California bay *Uc* FatB1 has a strong preference for 12:0-ACP (and a modest preference for 14:0-ACP; Voelker and Davies, 1994). A new *C. hookeriana* thioesterase, encoded by *Ch FatB2* (represented by SV2 in Figure 1; see Figure 4 for its position in the tree), has been characterized and found to hydrolyze 8:0- and 10:0-ACP (K. Dehesh, D. Knutzon, A. Jones, and T.A. Voelker, unpublished data). It is the enzyme involved in 8:0 and 10:0 fatty acid production for the storage lipids in seed. In addition, the specificity of a FatB representative from elm (*U. americana*) seed, *Ua* FatB1 (Figure 4; D. Knutzon, A. Jones, H.M. Davies, and T.A. Voelker, unpublished data), shows that this enzyme is involved in 10:0 production for the elm storage lipids (80 mol% 10:0; Davies, 1993). These two enzymes can therefore be considered functional homologs of California bay's *Uc* FatB1. In summary, FatB representatives with hydrolytic specificities matching the predominant medium-chain fatty acids of the respective seed oils have been found in species of three taxonomically dispersed angiosperm families (Lauraceae, Lythraceae, and Ulmaceae). Medium-chain ACP hydrolytic activity was also detected in maturing coconut endosperm (Davies, 1993). Therefore, it seems likely that medium-chain–preferring FatB representatives are common, if not universal, components of the fatty acid synthases specialized for medium-chain production in oilseeds.

The expression and enzymatic properties of this medium-chain–specific FatB subgroup contrast markedly with those of *Ch* FatB1 from *C. hookeriana* and *At* FatB1 from Arabidopsis. Both enzymes cover the C14 to C18 range, with 16:0-ACP preference. Also, expression of their genes was found not to be restricted to seed. To date the *FatB* cDNAs isolated from several other species (some of which do not accumulate medium-chain fatty acids in storage lipids) also encode thioesterases with this C14 to C18 specificity (D. Hawkins, J. Kridl, L. Yuan, and T.A. Voelker, unpublished data).

Thus, FatA and FatB also differ functionally. The FatA thioesterases appear invariably specific for one unsaturated substrate, namely, 18:1-ACP, irrespective of the organ or species from which they are isolated. In contrast, FatB members prefer saturated acyl-ACPs, and substrate chain length varies greatly from 8:0-ACP to 18:0-ACP. In addition, FatB members fall into two functional groups. Some FatB enzymes are specific for saturated ACPs in the C8 to C14 range and to date are found only in medium-chain–producing species, with expression restricted to medium-chain–producing tissues. Enzymes of a second FatB group preferring C14 to C18 acyl-ACPs (predominantly 16:0) are probably present in all major plant parts and are not restricted to medium-chain–producing species.

Is there prior evidence for a 16:0-ACP thioesterase in plants? Because it was possible to produce high-palmitate mutants of Arabidopsis, James and Dooner (1990) first proposed the

existence of a 16:0-ACP–hydrolyzing activity that was separate from the oleoyl thioesterase, FatA. Although plausible, this phenotype could also be caused by a mutation affecting β-ketoacyl-(ACP) synthase II (Browse and Somerville, 1991) that should diminish extension of the acyl chains beyond C16. The minor activity of FatA on 16:0-ACP (Knutzon et al., 1992) could have hydrolyzed the accumulating 16:0-ACP. By analyzing transformed, FatA cosuppressed somatic soybean embryos, Yadav et al. (1993) deduced from acyl-ACP hydrolysis measurements of these tissues that a fraction of the 16:0 and 18:0 production resisted suppression of the resident FatA. These authors interpreted this as evidence for an additional thioesterase. In addition, crude extracts of palm mesocarp, a tissue naturally producing high-palmitate oil, showed a marked elevation of 16:0-ACP hydrolysis compared with other plant tissues (Sambanthamurthi and Oo, 1990). This could be caused by the up regulation of a *Ch* FatB1–like enzyme, as we have demonstrated with transgenic canola (Figure 5). Indeed, the relation between 16:0/18:1–ACP hydrolysis in extracts of our high-palmitate *Brassica* seed resembles those measured with the palm mesocarp extracts.

Evolution of Plant Acyl-ACP Thioesterases

It follows, from its taxonomic distribution and expression pattern, that the C14 to C18 form of FatB is present in all organs (if not all cells) of higher plants and is therefore most likely the ancestral version of FatB. The 16:0-ACP–preferring FatB1 enzymes of *C. hookeriana* and Arabidopsis are examples of such normal FatB enzymes. We further conclude that in plants whose seed triglycerides contain medium chains, a FatB with medium-chain specificity evolved from the resident normal C14 to C18 FatB concomitantly with a change to seed-restricted expression of its gene. This secundary specialization of function and expression occurred independently several times during the evolution of fairly unrelated angiosperm families or genera (parallel evolution).

For example, the genus *Cuphea*, with fossil evidence only traced back to about 7 million years (Graham and Graham, 1971), is the only genus with medium chains in the Lythraceae (Graham and Kleiman, 1987). Our hypothesis implies that during this period, the gene for the medium-chain–preferring *Ch* FatB2 developed from a C14 to C18 thioesterase gene, such as *Ch FatB1.* On the other hand, because *Ch FatB1* and *Cl FatB1* are more similar to each other than either one is to *Ch FatB2* (Figure 4), the *Ch FatB2* progenitor must have separated from its ancestor before the divergence of the evolutionary lineages leading to the two species. Given their positions in the tree, the two genes appear orthologous (Fitch, 1970), and we predict that the enzymatically uncharacterized *Cl* FatB1 has an activity profile similar to that of *Ch* FatB1.

Such a correlation between overall primary sequence and specificity profile cannot be applied when comparing FatB enzymes from different families. For example, the 8:0/10:0-ACP thioesterase of *C. hookeriana* (*Ch* FatB2) is much more sequence related to the 16:0-ACP thioesterase of its own species (*Ch* FatB1) than to the 8:0/10:0-ACP thioesterase of elm (*Ua* FatB1) or the 12:0-ACP thioesterase of bay (*Uc* FatB1). Amino acid sequence alignments of all FatB representatives approximately replicated the DNA sequence–based tree of Figure 4 (data not shown). This is expected, because the change in specificity is most likely achieved by changing only very few amino acids, and many different solutions might be possible (as shown for lipoxygenases; Sloane et al., 1991).

The evolution and preservation of two distinct acyl-ACP thioesterases in plants raise three principal questions: (1) Was the specificity of the ancient "Fat" thioesterase closer to today's FatA or to FatB? (2) What necessitated the divergence of the original "Fat" into the two classes? (3) What are the current roles of FatA and FatB in the cellular machinery? Extensive searches in current sequence data bases failed to uncover any significant relationships of FatA/B with any other entry. Even the known non-plant acyl-ACP thioesterases (Lawson et al., 1994) appear to have a different origin. To date, therefore, no clues for the evolutionary source of Fat, which could have assisted with the evaluation of its original specificity, have been uncovered by this approach.

The ubiquity of 18:1 in many organisms suggests that 18:1 specificity (FatA-like) is ancient. Why then do plants need FatB? It is possible that the duplication of Fat was correlated with the emergence of terrestrial plants. The shift from an aquatic to a terrestrial habitat necessitated the development of diffusion barriers for water; these barriers in extant plants are made up largely of saturated fatty acid derivatives (reviewed by Wettstein-Knowles, 1993). It is conceivable that this new, increased requirement for saturated long-chain fatty acids was met by the development of FatB.

However, a comparison of the fatty acid synthases of plants and cyanobacteria suggests a different evolutionary scenario. It is currently widely accepted that the plastid organelles of plants are direct descendants of endosymbiotic cyanobacteria ("endosymbiont hypothesis"; reviewed in Gray, 1989). This is also reflected in the extensive structural and sequence similarities of the respective (type II) fatty acid synthases and other plant lipid synthetic enzymes (Nishida et al., 1993). However, acyl-ACP thioesterases could not be detected in cyanobacteria. The ACP-bound fatty acid synthesis is terminated exclusively by the "prokaryotic" mechanism, that is, by transferring the acyl-chains to a glycerol moiety (Nishida et al., 1993). Also, cyanobacteria lack the soluble 18:0-ACP desaturase common to plastids. Interestingly, this plant 18:0 desaturase is an enzyme that (like FatA/B) does not appear homologous to the functionally analogous enzymes of non-plant organisms (Browse and Somerville, 1991). Therefore, in contrast to plastids, desaturases of cyanobacteria act exclusively on lipids; the end products of the cyanobacterial fatty acid synthase are all saturated (Wada et al., 1993).

Assuming that these features of cyanobacterial fatty acid synthesis are ancient, we propose the following hypothesis

of Fat evolution: after engulfment of the cyanobacterial endosymbiont, the resulting plant cells possessed two fatty acid biosynthesis pathways. The unicellular plant *Euglena*, which carries prokaryotic and eukaryotic de novo fatty acid synthases (Siebenlist et al., 1991), could serve as a model for such an intermediate form. In a subsequent stage of interdependence, the early chloroplast came to play the role of the exclusive de novo fatty acid–producing cell compartment, and concurrently an acyl export system developed. The source from which the acyl-ACP thioesterase ("Fat") emerged remains obscure. If the source for Fat was chloroplastic, its gene subsequently must have been transferred to the nucleus. Alternatively, it is conceivable that a nuclear-encoded enzyme was recruited. In both scenarios a transit peptide had to develop.

Because the acyl-ACP–based fatty acid synthase of ancient chloroplasts probably supplied only saturated end products (the fatty acid synthases of cyanobacteria produce predominantly 16:0-ACP and some 18:0-ACP; Murata and Nishida, 1987), it is tempting to speculate that the ancient plastidial thioesterase, "Fat," had FatB-like specificity; that is, it preferred long-chain saturated acyl-ACPs, such as 16:0. As shown by our *Ch* FatB1 expression studies in *E. coli*, the 16:0-ACP thioesterase can hydrolyze 18:1-ACP (Figures 3 and 5A). It is conceivable that an 18:1-ACP thioesterase (FatA) subsequently evolved from "Fat" (with the advent of the soluble 18:1-ACP desaturase). The ancient "Fat" specificity was conserved in the FatB lineage; maybe the availability of two thioesterases with different specificities is advantageous to plants in order to maintain flexibility in the plastidial export of acyl groups. More comparative studies of acyl-ACP thioesterases of divergent plant lineages, such as lower vascular plants or algae, will enable scientists to trace further the development of the plastidial fatty acid export system, which is a crucial element in the lipid biosynthesis pathway of higher plants.

METHODS

Nomenclature for Plant Acyl–Acyl Carrier Protein Thioesterases

Currently, cDNA sequences for more than a dozen plant thioesterases are known, and they fall into two distinct sequence classes. We have introduced a mnemonic terminology according to the emerging terminology for plant genes and their gene products (Commission of Plant Gene Nomenclature, 1994) and have used the acronym *Fat* for plant fatty acyl–acyl carrier protein (ACP) thioesterase. The uppercase letters *A* and *B* designate the respective sequence subclass, as defined in a following section. The Arabic numeral defines a certain allele described for a species but carries no functional information. The mnemonic is preceded either by the full species name or by a two-letter abbreviation derived from the first letters of the genus and species. Using this terminology, the gene (or cDNA sequence) of the California bay (*Umbellularia californica*) medium-chain thioesterase is described as *Uc FatB1*, and its gene product is given as *Uc* FatB1.

Plant Material

Seeds of *Cuphea hookeriana* were obtained from S. Graham, Kent State University (Kent, OH; accession number 202, originally collected in Oaxaca, Mexico), and plants were grown in the greenhouse. Developing seeds were harvested during the phase of oil accumulation, frozen in liquid nitrogen, and stored at −70°C.

RNA Isolation

A protocol for the isolation of total RNA from developing seed of *C. hookeriana* was developed from a hexadecyltrimethylammonium bromide (CTAB) DNA isolation procedure (Webb and Knapp, 1990). Developing *C. hookeriana* seeds (1 g) were ground in liquid nitrogen with a mortar and pestle. The powder was homogenized with mortar and pestle at room temperature with 10 mL of extraction buffer (50 mM Tris-HCl, pH 9, 0.8 M NaCl, 10 mM EDTA, 0.5% [w/v] CTAB, 1% [v/v] β-mercaptoethanol, 2% [w/v] polyvinylpolypyrrolidone), and the insoluble fraction was sedimented for 5 min at 15,000*g*. The supernatant was filtered through Miracloth (Calbiochem) and deproteinized with 0.3 volumes of chloroform. After phase separation by centrifugation (3 min at 10,000*g*), 1.5 volumes of CTAB precipitation buffer (50 mM Tris-HCl, pH 9, 10 mM EDTA, 0.5% [w/v] CTAB, 0.1% [v/v] β-mercaptoethanol) was added. After a 30-min incubation at room temperature, the nucleic acids were sedimented by centrifugation (15,000*g* for 30 min). The pellet was dissolved in 0.4 mL of 1 M NaCl, 0.1% (v/v) β-mercaptoethanol. After extraction with 1 volume of phenol chloroform (1:1 [v/v]) and ethanol precipitation, the nucleic acids were redissolved in water. Typical yields were 0.5 to 1 mg of total RNA with only traces of genomic DNA. Total RNA was further purified by LiCl precipitation, and poly(A) RNA was isolated by two cycles of chromatography on oligo(dT) cellulose (Sambrook et al., 1989).

Generation of cDNA Probes by Polymerase Chain Reaction

Degenerate polymerase chain reaction (PCR) primers were designed for amino acid sequences conserved among *Ct* FatA1, *Ct* FatA2 (Knutzon et al., 1992), and *Uc* FatB1 (Voelker et al., 1992). To reduce degeneracy, three- and fourfold degenerate base positions were substituted by a 1:1 deoxyinosine/deoxycytosine mix. Deoxyinosine was added to the 3′ ends to improve pairing of primer 3′ ends with the template (Batzer et al., 1991). Restriction enzyme sites were also added to the 5′ ends to facilitate cloning of PCR products. The oligonucleotides were synthesized on an ABI 380A DNA synthesizer (Applied Biosystems, Foster City, CA). Primers of the successful combinations (see Figure 1 for anchor locations; degenerate positions are in brackets; I denotes inosine) are as follows: S-sense, 5′-GCGGATCCTA(C/T)-CC(I/C)RC(I/C)TGG(A/G)T)(C/G)(I/G)GA-3′; V-antisense, 5′-GCCTCGAG-(A/G)TT(A/G)TT(ACGT)AC(A/G)TG(C/T)T(G/T)(A/G)TTI-3′; V-sense, 5′-GCGGATCCAA(C/T)(C/A)A(A/G)CA(C/T)GT(ACGT)AA(C/T)AA(C/T)-I-3′; X-antisense, 5′-GCGGATCCGA(I/C)TA(C/T)(C/A)G(I/C)(C/A)G(I/C)-GA(A/G)TG(C/T)I-3′. Single-stranded cDNA template for reverse transcriptase–PCR was synthesized at 45°C from poly(A) RNA with Superscript (Gibco BRL). A 100-μL PCR (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% [w/v] gelatin, 200 μM of each deoxynucleotide triphosphate) contained 100 pmol of each primer and single-stranded cDNA derived from 40 ng of poly(A) RNA. Before PCR, the mixture was heated to 100°C for 5 min and cooled to 70°C; 5 U

of Amplitaq DNA polymerase (Perkin-Elmer Cetus) were then added. Thioesterase sequences were amplified in a Perkin-Elmer Cetus DNA thermal cycler during 40 cycles of the following program: 94°C for 60 sec, 65°C for 1 sec, a 120-sec ramp from 65 to 40°C, 40°C for 30 sec, and 72°C for 60 sec.

cDNA Library Construction and Screening

The cDNA was synthesized according to Stratagene's ZAP cDNA synthesis kit using 5 μg of poly(A) RNA and ligated to either a pBluescript II KS− plasmid or the bacteriophage Lambda UNI-ZAP XR (Stratagene). Plasmid libraries were screened by replicating colonies onto nitrocellulose filters and using ^{32}P-labeled DNA probes (Sambrook et al., 1989). Hybridization was at 42°C for 20 hr in 50% (v/v) formamide, 5 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 50 mM sodium phosphate, pH 7, 5 × Denhardt's solution (1 × Denhardt's solution is 0.02% Ficoll, 0.02% PVP, 0.02% BSA), 0.1% (w/v) SDS, 5 mM EDTA, pH 8, 10 μg/mL denatured salmon sperm DNA (Sambrook et al., 1989). Filters were washed in 0.1 × SSC, 0.1% (w/v) SDS at 55°C. The bacteriophage λ library was screened under less stringent conditions (formamide concentration 30% [v/v]), and filters were washed in 1 × SSC, 0.1% (w/v) SDS at room temperature. Cloned cDNAs were sequenced according to the method of Hattori and Sakaki (1986). Sequencing was done either manually using ^{35}S-dCTP or with an automated ABI 373A sequencer (Applied Biosystems).

Sequence Handling

Sequences were analyzed on VAX (Digital Equipment Corp., Woburn, MA) and Macintosh (Apple Computer Inc., Cupertino, CA) computers. Before cladistic analysis, cDNA sequences were aligned using the "region method" of the Genalign program of the Intelligenetics Suite, version 5.3 (Intelligenetics Inc., Mountain View, CA). The multiple alignment was further adjusted manually and used for the calculation of a most parsimonious (shortest) tree using the branch-and-bound algorithm of the Macintosh version of PAUP, release 3.0 (Swofford, 1991). The program was executed with the following settings: "Gapmode=missing"; "Addition: Furthest"; "Collapse zero branches." The tree was rooted at the longest branch ("Midpoint Rooting"). Dot matrix alignments were executed using the Pustell's dot matrix algorithm, as modified for the McVector program, version 4.1 (IBI, New Haven, CT).

Construction of LacZ-Thioesterase Fusion Protein Expression Vectors

The *Ch FatB1* cDNA clone pCGN4806 was digested with StuI and XhoI, and the 1.2-kb DNA fragment was cloned into pBC SK (Stratagene), which itself was cut with ClaI (subsequently filled in to create a blunt end) and XhoI, resulting in pCGN4811. The resulting 339-amino acid fusion protein contained 296 amino acids of *Ch* FatB1 from residue 119 to 416 (Figure 1).

Plant Transformation

For seed-specific expression in plants, we engineered a SalI site 6 bp 5′ of the proposed translational start and an XhoI site 171 bp 3′ of the stop codon. The engineered *Ch FatB1* reading frame was inserted into pCGN3223 (Kridl et al., 1991). This plasmid supplied a seed-specific napin promoter and a napin 3′ termination fragment from *Brassica rapa*. The chimeric gene was inserted as a HindIII fragment into the binary plant transformation vector pCGN1578 (McBride and Summerfelt, 1990) resulting in pCGN3800, which was used to transform *B. napus* cv 212/86 (Radke et al., 1988). Transgenic plants were regenerated and then grown in a greenhouse.

Enzyme Activity Assays and Lipid Analysis

For activity determinations with *Escherichia coli* lysates, the strain DH5α (Gibco BRL; a derivative of DH5; Hanahan, 1985) was transformed with pCGN4811 and with a control vector. Cultures were grown at 30°C to an OD_{600} of 0.4 and induced with isopropyl β-D-thiogalactopyranoside to 0.4 mM for 2 hr. The cells were sedimented by centrifugation, resuspended in thioesterase assay buffer (Voelker et al., 1992), and lysed by one freeze/thaw cycle and sonication. Debris was sedimented by a 5-min centrifugation at 14,000*g*, and supernatants were assayed as described previously (Voelker and Davies, 1994). For bacterial lipid analysis, a fatty acid degradation-minus *E. coli* strain, K27 (*fadD88*) (Klein et al., 1971), was transformed with the respective expression plasmids. Cultures were grown in Luria-Bertani broth at 30°C. Total culture lipids were extracted and analyzed (Voelker and Davies, 1994), and free fatty acids were separated from phospholipids by thin-layer chromatography (Cho and Cronan, 1994). Enzyme and fatty-acyl composition analyses of *B. napus* seed were as described previously (Browse et al., 1986; Voelker et al., 1992).

ACKNOWLEDGMENTS

We are indebted to Calgene's *Brassica* transformation, growth chamber, greenhouse, and lipid analysis personnel, who supported us with their skills and time. Sheryl Eriqat and Kristine Michalik helped with assays, Ann Cranmer and Katie Dehesh made the cDNA libraries, and Debbie Knutzon isolated the longest cDNA clone. We thank Katie Dehesh, Debbie Hawkins, Jean Kridl, and Ling Yuan (Calgene Inc.) and Peter Dörmann and John Ohlrogge (Michigan State University, East Lansing) for making unpublished data available to us. Mark Chase and Jeffrey Palmer sent us a manuscript prior to publication. Scientists at Calgene and John Ohlrogge contributed in discussions and critically read the manuscript.

Received January 4, 1995; accepted January 23, 1995.

REFERENCES

Batzer, M.A., Carlton, J.E., and Deininger, P. (1991). Enhanced evolutionary PCR using oligonucleotides with inosine at the 3′-terminus. Nucl. Acids Res. **19,** 5081.

Browse, J., and Somerville, C.R. (1991). Glycerolipid synthesis: Biochemistry and regulation. Annu. Rev. Plant Physiol. Plant Mol. Biol. **42,** 467–506.

Browse, J., McCourt, P.J., and Somerville, C.R. (1986). Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal. Biochem. **152,** 141–145.

Bubeck, D.M., Fehr, W.R., and Hammond, E.G. (1989). Inheritance of palmitic and stearic acid mutants of soybean. Crop Sci. **29,** 652–656.

Chase, M.W., et al. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. **80,** 528–580.

Cho, H., and Cronan, J.E., Jr. (1994). "Protease I" of *Escherichia coli* functions as a thioesterase *in vivo.* J. Bacteriol. **176,** 1793–1795.

Commission on Plant Gene Nomenclature (1994). A nomenclature for sequenced plant genes: A brief description. Plant Mol. Biol. Rep. **12**(suppl.), S4–S9.

Davies, H.M. (1993). Medium chain acyl-ACP hydrolysis activities of developing oilseeds. Phytochemistry **33,** 1353–1356.

Davies, H.M., Anderson, L., Fan, C., and Hawkins, D.J. (1991). Developmental induction, purification, and further characterization of 12:0-ACP thioesterase from immature cotyledons of *Umbellularia californica.* Arch. Biochem. Biophys. **290,** 37–45.

de Boer, A.D., and Weisbeek, P.J. (1991). Chloroplast protein topogenesis: Import, sorting and assembly. Biochim. Biophys. Acta **1071,** 221–253.

Dörmann, P., Spener, F., and Ohlrogge, J.B. (1993). Characterization of two acyl-acyl carrier protein thioesterases from developing *Cuphea* seeds specific for medium-chain and oleoyl-acyl carrier protein. Planta **189,** 425–432.

Dörmann, P., Kridl, J., and Ohlrogge, J.B. (1994). Cloning and expression in *Escherichia coli* of a cDNA coding for the oleoyl-acyl carrier protein thioesterase from coriander (*Coriandrum sativum* L.). Biochim. Biophys. Acta **1212,** 134–136.

Dörmann, P., Voelker, T.A., and Ohlrogge, J.B. (1995). Cloning and expression in *E. coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long chain acyl-acyl carrier proteins. Arch. Biochem. Biophys. Res. Commun. **316,** 612–618.

Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. Syst. Zool. **19,** 99–113.

Gibson, S., Falcone, D.L., Browse, J., and Somerville, C.R. (1994). Use of transgenic plants and mutants to study the regulation and function of lipid composition. Plant Cell Environ. **17,** 627–637.

Gould, S., Subramani, S., and Scheffler, I.E. (1989). Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. Proc. Natl. Acad. Sci. USA **86,** 1934–1938.

Graham, A., and Graham, S. (1971). Geologic history of the Lythraceae. Brittonia **23,** 335–346.

Graham, S.A., and Kleiman, R. (1987). Seed lipids of the Lythraceae. Biochem. Syst. Ecol. **15,** 433–439.

Graham, S.A., Hirsinger, F., and Röbbelen, G. (1981). Fatty acids of *Cuphea* (Lythraceae) seed lipids and their systematic significance. Am. J. Bot. **68,** 908–917.

Gray, M. (1989). The evolutionary origins of organelles. Trends Genet. **5,** 294–299.

Grellet, F., Cooke, R., Raynal, M., Laudié, M., and Delseny, M. (1993). *Arabidopsis thaliana* systematic cDNA sequencing reveals a gene with homology with *Umbellularia californica* C12:0-ACP thioesterase. Plant Physiol. Biochem. **31,** 599–602.

Hanahan, D. (1985). *E. coli* DH5α. In Cloning: A Practical Approach, Vol. I, D.M. Glover, ed (Oxford: IRL Press), pp. 109–114.

Hattori, M., and Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152,** 232–237.

Higgins, D.G., and Sharp, P.M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. Comp. Appl. Bio. Sci. **5,** 151–153.

Hilditch, T.P., and Williams, P.N. (1964). The Chemical Constitution of Natural Fats. (New York: John Wiley & Sons).

Hitz, W.D., and Yadav, N.S. (1992). Nucleotide sequences of soybean acyl-ACP thioesterase genes. International Patent No. WO 92/11373. Issued July 9, 1992.

James, D.W., Jr., and Dooner, H.K. (1990). Isolation of EMS-induced mutants in *Arabidopsis* altered in seed fatty acid composition. Theor. Appl. Genet. **80,** 241–245.

Klein, K., Steinberg, R., Fiethen, B., and Overath, B. (1971). Fatty acid degradation in *Escherichia coli*: An inducible system for the uptake of fatty acids and further characterization of old mutants. Eur. J. Biochem. **19,** 442–450.

Knutzon, D.S., Bleibaum, J.L., Nelsen, J., Kridl, J.C., and Thompson, G.A. (1992). Isolation and characterization of two safflower oleoyl-acyl carrier protein thioesterase cDNA clones. Plant Physiol. **100,** 1751–1758.

Kridl, J.C., McCarter, D.W., Rose, R.E., Scherer, D.E., Knutzon, D.S., Radke, S.E., and Knauf, V.C. (1991). Isolation and characterization of an expressed napin gene from *Brassica rapa.* Seed Sci. Res. **1,** 209–219.

Lawson, D.M., Derewenda, U., Serre, L., Ferri, S., Szittner, R., Wei, Y., Meighen, E.A., and Derewenda, Z.S. (1994). Structure of a myristoyl-ACP-specific thioesterase from *Vibrio harveyi.* Biochemistry **33,** 9382–9388.

Loader, N.M., Woolner, E.M., Hellyer, A., Slabas, A.R., and Safford, R. (1993). Isolation and characterization of two *Brassica napus* embryo acyl-ACP thioesterase cDNA clones. Plant Mol. Biol. **23,** 769–778.

Löhden, I., and Frentzen, M. (1988). Role of plastidial acyl-acyl carrier protein: Glycerol 3-phosphate acyltransferase and acyl-acyl carrier protein hydrolase in channeling the acyl flux through the prokaryotic and eucaryotic pathway. Planta **176,** 506–512.

Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. (1987). Selection of AUG initiation codons differs in plants and animals. EMBO J. **6,** 43–48.

Magnuson, K., Jackowski, S., Rock, C.O., and Cronan, J.E. (1993). Regulation of fatty acid biosynthesis in *Escherichia coli.* Microbiol. Rev. **57,** 522–542.

Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H., and Cerff, R. (1993). Molecular phylogenies in angiosperm evolution. Mol. Biol. Evol. **10,** 140–162.

McBride, K.E., and Summerfelt, K.R. (1990). Improved binary vectors for *Agrobacterium*-mediated plant transformation. Plant Mol. Biol. **14,** 269–276.

Morden, C.W., and Golden, S.S. (1989). *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature **337,** 382–384.

Murata, N., and Nishida, I. (1987). Lipids of blue-green algae (cyanobacteria). In The Biochemistry of Plants, Vol. 9, P.K. Stumpf, ed. (Orlando, FL: Academic Press), pp. 315–347.

Nishida, I., Imai, H., Ishizaki-Nishizawa, O., Tasaka, Y., Shiraishi, H., Higashi, S., Hayashi, H., Beppu, T., Matsuo, T., and Murata, N. (1993). Molecular and physiological studies of glycerol-3-phosphate acyltransferase, acyl-ACP hydrolase and stearoyl-ACP desaturase. In Biochemistry and Molecular Biology of Membrane and Storage Lipids of Plants, N. Murata and C.R. Somerville, eds (Rockville, MD: The American Society of Plant Physiologists), pp. 79–88.

Persson, C. (1985). High palmitic content in summer turnip rape (*Brassica campestris* var. *annua* L.). Cruciferae Newslett. Scott. Crop Res. Inst. **10,** 137.

Pollard, M.R., Anderson, L., Fan, C., Hawkins, D.H., and Davies, H.M. (1991). A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of *Umbellularia californica.* Arch. Biochem. Biophys. **284,** 306–312.

Radke, S.E., Andrews, B.M., Moloney, M.M., Crouch, M.L., Kridl, J.C., and Knauf, V.C. (1988). Transformation of *Brassica napus* L. using *Agrobacterium tumefaciens*: Developmentally regulated expression of a reintroduced napin gene. Theor. Appl. Genet. **75,** 685–694.

Safford, R. (1993). Opportunities for genetic engineering of oil production. In Seed Storage Compounds, P.R. Shewry and K. Stobart, eds (Oxford, UK: Clarendon Press), pp. 343–358.

Sambanthamurthi, R., and Oo, K.-C. (1990). Thioesterase activity in the oil palm (*Eloeis guineensis*) mesocarp. In Plant Lipid Biochemistry, Structure and Utilization, P.J. Quinn and J.L. Harwood, eds (London: Portland Press Ltd.), pp. 166–168.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sanderson, M.J., and Donoghue, M.J. (1994). Shifts in diversification rate with the origin of angiosperms. Science **264,** 1590–1593.

Siebenlist, U., Wohlgemuth, S., Finger, K., and Schweizer, E. (1991). Isolation of a novel type-I fatty acid synthetase from *Euglena gracilis.* Eur. J. Biochem. **202,** 515–519.

Sloane, D.L., Leung, R., Craik, C.S., and Sigal, E. (1991). A primary determinant for lipoxygenase positional specificity. Nature **354,** 149–152.

Swofford, D.L. (1991). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0r. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Töpfer, R., and Martini, N. (1994). Molecular cloning of cDNAs or genes encoding proteins involved in *de novo* fatty acid biosynthesis in plants. J. Plant Physiol. **143,** 416–425.

Voelker, T.A., and Davies, H.M. (1994). Alteration of the specificity and regulation of fatty acid synthesis of *E. coli* by expression of a plant medium-chain acyl-ACP thioesterase. J. Bacteriol. **176,** 7320–7327.

Voelker, T.A., Worrell, A.M., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D.H., Radke, S.E., and Davies, H.M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science **257,** 72–74.

Wada, H., Gombos, Z., Sakamoto, T., Higashi, S., Los, D.A., Heinz, E., Schmidt, H., Nishida, I., and Murata, N. (1993). Fatty acid desaturation in cyanobacteria. In Biochemistry and Molecular Biology of Membrane and Storage Lipids of Plants, N. Murata and C.R. Somerville, eds (Rockville, MD: The American Society of Plant Physiologists), pp. 67–78.

Webb, D.M., and Knapp, S. (1990). DNA extraction from a previously recalcitrant plant genus. Plant Mol. Biol. Rep. **8,** 180–185.

Wettstein-Knowles, P.M. (1993). Waxes, cutin, and suberin. In Lipid Metabolism in Plants, T.S. Moore, ed (Boca Raton, FL: CRC Press), pp. 127–166.

Yadav, N., Wierzbicki, A., Knowlton, S., Pierce, J., Ripp, K., Hitz, W., Aegerter, M., and Browse, J. (1993). Genetic manipulation to alter fatty acid profiles of oilseed crops. In Biochemistry and Molecular Biology of Membrane and Storage Lipids of Plants, N. Murata and C.R. Somerville, eds (Rockville, MD: The American Society of Plant Physiologists), pp. 60–66.