This is to certify that the

dissertation entitled

RICINOLEIC ACID BIOSYNTHESIS IN RICINUS COMMUNIS

presented by

Frank Joost van de Loo

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Botany & Plant Pathology

Major professor

Date 3/11/94

RICINOLEIC ACID BIOSYNTHESIS IN *RICINUS COMMUNIS*

By

Frank Joost van de Loo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Botany and Plant Pathology
and
Department of Energy Plant Research Laboratory

1993

ABSTRACT

RICINOLEIC ACID BIOSYNTHESIS IN *RICINUS COMMUNIS*

By

Frank Joost van de Loo

Ricinoleic acid, the hydroxylated fatty acid synthesized in developing seeds of castor (*Ricinus communis* L.) is the industrially-important constituent of castor oil. Little biochemical and no molecular information was available concerning the nature of the oleate-12-hydroxylase responsible for this unusual fatty acid. Characterisation of this enzyme and isolation of the encoding gene(s) were the goal of this work, toward the eventual aim of engineering ricinoleic acid production in an alternative oilseed.

The induction of hydroxylase activity concomitant with differentiation of the cellular endosperm during seed ontogeny was characterised. Anti-cytochrome b_5 antibodies were used to demonstrate that cytochrome b_5 is the electron donor to the hydroxylase. That molecular oxygen is the source of the hydroxyl oxygen was demonstrated using $^{18}O_2$ and mass spectrometry.

Intractability of the hydroxylase to purification prompted genetic and molecular approaches to cloning the hydroxylase gene. Of 5300 yeast clones expressing individual cDNAs from developing castor endosperm, none showed altered fatty acid composition when analysed by gas chromatography. Hypothesizing that the hydroxylase is homologous to microsomal fatty acid desaturases, a clone of the *Brassica napus* Fad3 desaturase was used to isolate hybridising castor cDNAs. A

gene isolated by this approach was identified as a different desaturase (Fad7). Fad7 was also isolated from castor seed mRNA by PCR using a conserved desaturase sequence motif, but no putative hydroxylase clone was identified by these approaches.

Expressed Sequence Tags were generated from 468 moderately-abundant seed-specific cDNA clones of developing castor endosperm. Of these, 213 could be identified (Blastx score $\geq$ 80), including the identification (by homology to desaturases) of two putative clones encoding the hydroxylase. An analysis of the possible role of the identified clones in castor seed metabolism is presented. The putative hydroxylase gene (pFL2) showed features expected of this gene: strong, seed-specific expression, and ca. 37% amino acid sequence identity with membrane-bound desaturase genes. Expression of pFL2 in transgenic yeast and plants did not result in detectable accumulation of ricinoleic acid, nor other changes in fatty acid composition.

Based on this work, experiments can be designed to test the identity of pFL2 as the oleate-12-hydroxylase gene.

"To seek it with thimbles, to seek it with care;
To pursue it with forks and hope;
To threaten its life with a railway-share;
To charm it with smiles and soap!

"For the Snark's a peculiar creature, that won't
Be caught in a commonplace way.
Do all that you know, and try all that you don't:
Not a chance must be wasted today!"

ACKNOWLEDGEMENTS

I warmly thank Dr. Chris Somerville for sharing his enthusiasm, knowledge and wisdom, and for his support and confidence in my abilities. I also thank Drs John Ohlrogge, Pam Green and Mike Thomashow for their guidance.

I thank all members of the lab for their help and assistance at every turn, as well as other members of the Plant Research Laboratory.

I thank Dr Brian Fox for helpful discussions on oxygenase chemistry. Svieta Ndibongo, and Susan Lootens of the PRL Plant Biochemistry Facility, gave patient and competent technical assistance to the work described in chapter 5.

I am particularly grateful to Vlada Orbović and Ed Cahoon, Scott Peck and Antje Heese-Peck, Gary Vercruysse and Catherine Tamareille for their hospitality, not to mention their friendship.

Finally I thank the people of the United States of America for their commitment to the research enterprise and their generosity to those such as me who profit so splendidly from our involvement in it, and from our sojourn on this soil.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

18:1	Standard fatty acid abbreviations are used. These are of the format (number of carbon atoms in linear acyl chain):(number of double bonds). Position of unsaturation may be referred to by numbering from the carboxyl end (e.g. Δ9,12-18:2 for linoleic acid), or from the methyl end (e.g ω-9,ω-6-18:2).
IgG	Immunoglobulin
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SDS	Sodium dodecyl sulphate
FAME	Fatty acid methyl ester
GC	Gas chromatograph
TMS	Trimethylsilyloxy
EI	Electron impact
MS	Mass spectrometry

CHAPTER 1

INTRODUCTION

UNUSUAL FATTY ACIDS OF PLANTS

Extensive surveys of the fatty acid composition of seed oils from different species of higher plants have resulted in the identification of more than 210 naturally occuring fatty acids which can be broadly classified into one of eighteen structural classes (Table 1). The classes are defined by the number and arrangement of double or triple bonds and various functional groups, such as hydroxyls, ketones, epoxys, cyclopentenyl or cyclopropyl groups, furans or halogens. This level of structural diversity is similar to that of some of the least diverse families of plant secondary metabolites, a class of compounds which have been estimated to contain as many as 100,000 different structures.[3] A summary of the range of structures to be found in plant fatty acids, and a small number of examples of representative fatty acids, is presented in Table 2. Extensive lists of the amounts and sources of plant fatty acids are available in earlier reviews.[1,2,7-11]

The most commonly occurring fatty acids, which may occur in both membrane and storage lipids, are a small family of 16- and 18-carbon fatty acids which may have from zero to three, methylene-interrupted, *cis* unsaturations.

Table 1. Minimum summary of the kinds of fatty acids found in plants. The numbers were compiled by counting the numbers of distinct fatty acids described in compilations.[1,2]

Type	Number of Structures
saturated	14
monounsaturated	11
diunsaturated[a]	9
triunsaturated[a]	9
tetraunsaturated[a]	6
pentaunsaturated[a]	3
hexaunsaturated[a]	1
nonconjugated ethylenic[b]	32
conjugated ethylenic	15
acetylenic	26
monohydroxy	33
polyhydroxy	12
keto	5
epoxy	9
cyclopentenyl	15
cyclopropanoid	6
furanoid	2
halogenated	2

[a]all double bonds are methylene-interrupted *cis*-isomers
[b]configurations other than methylene-interrupted *cis*-double bonds

Table 2. Diversity of unusual plant fatty acids[a].

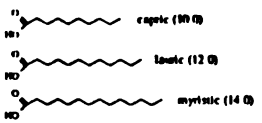
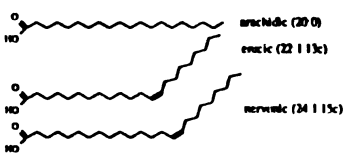
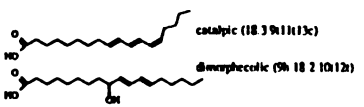
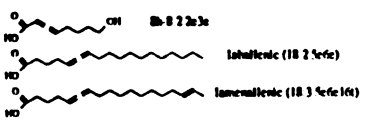
Unusual property	Name and structure	Source[b] (seed oil unless otherwise noted)
A. Chain length 1. Medium chain		Lauric is found extensively in Lauraceae and myristic extensively in Myristaceae. Commercially, medium-chain acids are obtained from coconut oil (*Cocos nucifera*, Arecaceae) and palm kernel oil (*Elaeis guineensis*, Arecaceae) and potentially from *Cuphea* spp. (Lythraceae).
2. Very long chain		Very long-chain acids occur in the Fabaceae and Sapindaceae (e.g., arachidic acid, 20:0, in *Nephelium lappaceum* and *Schleichera trifuga*; behenic acid, 22:0, in *Lophira* spp.) and also as wax esters in jojoba (*Simmondsia chinensis*, Simmondsiaceae). The most important commercial acid is erucic, common in the Brassicaceae, esp. rapeseed, mustard, and *Crambe abyssinica*. Potential for *Limnanthes* spp. (Limnanthaceae).
B. Unsaturation 1. *Trans*-ethylenic		Catalpic acid from *Catalpa ovata* (Bignoniaceae) and dimorphecolic acid from *Dimorphotheca aurantiaca* (Asteraceae).
2. Allenic		8h-8:2 in stillingia oil from kernels of Chinese Tallow Tree, *Stillingia sebifera* (syn. *Sapium sebiferum*; Euphorbiaceae). Laballenic acid from *Leonotis nepetaefolia* (Lamiaceae) and lamenallenic acid from *Lamium purpureum* (Lamiaceae)

Table 2 (cont'd).

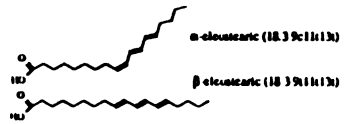
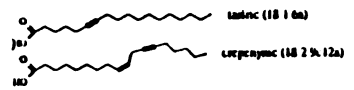
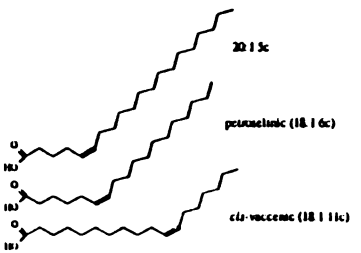
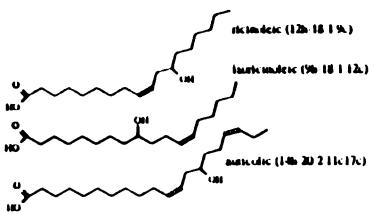
3. Conjugated		Commercially in tung oil, *Aleurites fordii* (Euphorbiaceae). Various other examples, generally also containing *trans* double bonds (e.g., catalpic acid, above). Industrially, conjugated diene (9,11) acids obtained by dehydration of ricinoleic acid from castor oil, yielding also nonconjugated diene (9,12) acids.
4. Acetylenic		Tariric acid from *Picramnia* spp. (Simaroubaceae) and crepenynic acid from *Crepis* spp. (Asteraceae). Various more complex acetylenic acids occur, often in the Olacaceae and Santalaceae.
5. Unusual positions		20:1 5c from *Limnanthes douglasii*, *L. alba* (Limnanthaceae). Petroselinic acid widespread in the Apiaceae and Araliaceae, e.g., coriander, carrot. *Cis*-vaccenic acid a relatively common minor acid but at higher levels in some seed oils (e.g., *Entandraphragma* spp., Meliaceae[4]) and tropical fruits (e.g., *Diospyros kaki*[5]). Various other examples, including erucic acid, above.
C. Oxygenated 1. Hydroxy		Ricinoleic acid from castor oil, *Ricinus communis* (Euphorbiaceae); isoricinoleic acid from *Strophanthus* spp., *Wrightia* spp. (Apocynaceae), and auricolic acid from *Lesquerella auriculata* (Brassicaceae). Some of these acids occur in various other species, castor oil is the most important commercial source. ω and ω-1 hydroxy acids are common components of

Table 2 (cont'd).

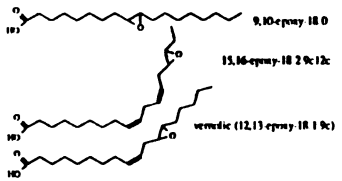
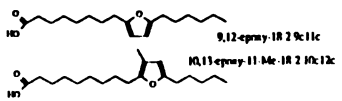
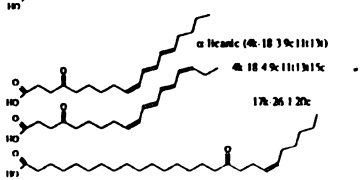
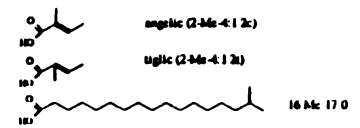
2. Epoxy		9,10 epoxy-18:0 from *Tragopogon porrifolius* (Asteraceae); vernolic acid from *Vernonia anthelmintica* (Asteraceae), *Euphorbia lagascae* (Euphorbiaceae); and 15,16 epoxy-18:2 from *Camelina sativa* (Brassicaceae).
3. Furanoid		9,12-epoxy-18:2 from *Exocarpus cupressiformis* (Santalaceae) and 10,13-epoxy-11-Me-18:2, a major component of rubber latex triacylglycerol (*Hevea brasiliensis*, Euphorbiaceae).
4. Keto		α-Licanic, a major fatty acid of oiticica oil (*Licania rigida*, Rosaceae); 4k-18:4 with other keto acids in *Chrysobalanus icaco* (Chrysobalanaceae); and 17k-26:1 with other keto acids in *Cuspidaria pterocarpa* (Bignoniaceae).

Table 2 (cont'd).

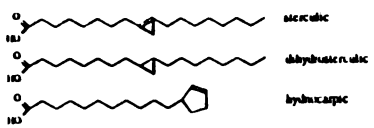
D. Other

1. Branched chain



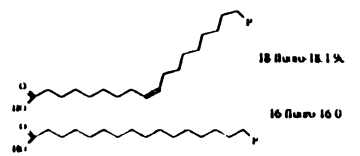
Angelic acid from roots of *Angelica archangelica* (Apiaceae) and tiglic acid from roots and seeds of *Croton tiglium* (Euphorbiaceae). 16Me-17:0 with other branched acids in vegetative and generative tissues of plastome mutants of *Antirrhinum majus* (Scrophulariaceae).[b] Branched chain acids are rare in plants and are much more common in bacteria.

2. Cyclic



Sterculic acid from *Sterculia foetida* (Sterculiaceae); dihydrosterculic acid from *Euphoria longana* (Sapindaceae); and hydnocarpic acid from chaulmoogra oil, *Hydnocarpus wightiana* (Flacourtiaceae). Cyclopentenyl acids in chaulmoogra oil formerly used in the treatment of leprosy.

3. Fluoro



Very rare acids found in seeds of *Dichapetalum toxicarium* (Dichapetalaceae).

[a] This table is intended to give an impression of the diversity of plant fatty acid structures, with a small number of examples of both the fatty acids and their sources only. Reviews cited in the text give a comprehensive coverage of the occurrence of unusual fatty acids.

[b] Compiled chiefly from References 1 and 2. Original references are given where not available from these sources.

[c] Abbreviations: c — *cis* ethylenic; t — *trans* ethylenic; e — ethylenic; a — acetylenic; h — hydroxy; k — keto, and Me — methyl.

All members of the family are descended from the fully saturated species as the result of a series of sequential desaturations which begin at the Δ9 carbon and progress in the direction of the Δ15 carbon.[12] Fatty acids which cannot be described by this simple algorithm are generally considered "unusual" even though several, such as lauric (12:0), erucic (22:1) and ricinoleic (12-OH, 18:1) are of significant commercial importance. The biosynthesis of ricinoleic acid in castor (*Ricinus communis*) seed is the focus of this work.

Much of the research to date concerning unusual plant fatty acids has been focused on the identification of new structures or cataloguing the composition of fatty acids found in various plant species. Relatively little is known about the mechanisms responsible for the synthesis and accumulation of unusual fatty acids, or of their significance to the fitness of the plants which accumulate them. This chapter gives a general review of unusual fatty acids in plants and a review of the data pertaining to biosynthesis of ricinoleic acid available when this work was commenced. Finally the goals of this work are summarised.

TAXONOMIC RELATIONSHIPS AMONG PLANTS ACCUMULATING UNUSUAL FATTY ACIDS

The taxonomic relationships between plants having similar or identical kinds of unusual fatty acids have been examined.[1,7] In some cases, particular fatty acids occur mostly or solely in related taxa. For example, the cyclopentenyl fatty acids have been

found only in the family Flacourtiaceae, although the presence of cyclopentenylglycine, the biosynthetic precursor of the cyclopentenyl fatty acids, in the Passifloraceae and Turneraceae suggests that these acids may also be found in these other families of the order Violales.[13,14] Petroselinic acid is most commonly found in the related families Apiaceae, Araliaceae and Garryaceae, but has also been observed in unrelated families.[15] In other cases there does not appear to be a direct link between taxonomic relationships and the occurrence of unusual fatty acids. For example, lauric acid is prominent in both the unrelated families Lauraceae and Arecaceae. Similarly, ricinoleic acid has now been identified in 12 genera from 10 families.[16-27] However in the large genus *Linum*, ricinoleic acid was found in the seed oil of all species tested from the section *Syllinum* but in no species from other sections of the genus.[17] If any conclusion can be drawn from the taxonomic distribution of unusual fatty acids, it would seem to be that the ability to synthesize some unusual fatty acids appears to have evolved several times independently, while for others it may have evolved only once.

TISSUE LOCALISATION OF UNUSUAL FATTY ACIDS

A feature of unusual fatty acids is that they are generally confined to seed triacylglycerols.[28-33] Analyses of vegetative tissues have generated few reports of unusual fatty acids, other than those occurring in the cuticle. A small number of exceptions exist in which unusual fatty acids are found in tissues other than the seed.

The Δ^6-desaturated acids γ-linolenic ($\Delta^{6,9,12}$-18:3) and octadecatetraenoic ($\Delta^{6,9,12,15}$-18:4) are found in leaves and seeds of some plants of the Boraginaceae.[34,35] The cyclopropenoid fatty acids of the order Malvales are not restricted to the seeds.[36-38] The cyclopentenyl fatty acids of Flacourtiaceae seed oils are also found in smaller proportions in the phospholipids and glycolipids of various tissues.[39] Acetylenic fatty acids are found in root, stem and leaf in the Santalaceae.[11,36] The small triacylglycerol fraction of rape leaves and siliques contains erucic acid.[36] Similarly, petroselinic acid is found in the pericarp (a maternal tissue) as well as within the seeds of ivy (*Hedera helix*).[36] Latex of the rubber tree, *Hevea brasiliensis*, contains triacylglycerol in which a furanoid acid (Table 2) is the major component,[40] while seeds of the same tree contain no unusual fatty acids.[41] An interesting observation is that branched chain fatty acids accumulate in the yellow-white chloroplast-deficient parts of leaves of plastome mutants of snapdragon and tobacco, which have reduced levels of unsaturated fatty acids. The branched chain fatty acids are completely absent from normal leaves and normal green parts of mosaic leaves.[6]

POTENTIAL PHYSIOLOGICAL ROLES OF UNUSUAL FATTY ACIDS IN PLANTS

Since the ability to synthesize various unusual fatty acids must have evolved independently, the common feature of confinement to seed triacylglycerol indicates some selective constraint or functional significance. One possible function of unusual

fatty acids is that by being toxic or indigestible they protect the seed against herbivory. Some unusual fatty acids may be inherently toxic, such as the acetylenic fatty acids, or some of their metabolites described below, which have antibiotic properties.[11] Other unusual fatty acids are toxic upon catabolism by the herbivore, such as the ω-fluoro fatty acids of *Dichapetalum toxicarium*.[42] One of these acids (*threo*-18-fluoro-9,10-dihydroxystearic acid) was lethal to rats when injected intraperitoneally at 25 mg kg^{-1}, probably due to catabolism to toxic fluoroacetate. The cyclopentenyl fatty acids were long used in the treatment of leprosy, and activity of hydnocarpic acid (Table 2) against many *Mycobacterium* species has been demonstrated.[39,43] These acids were also deleterious to the patients, causing a range of side-effects. The cyclopropenoid fatty acids also appear to have biological activities, possibly due to the accumulation in animal tissues of partial catabolites containing the cyclopropene ring, which inhibits β-oxidation.[39] Three effects of cyclopropenoid fatty acids have been described but are poorly understood: there is some alteration of the properties of membranes, there is an inhibition of fatty acid desaturase activity, and there is a carcinogenic effect, or a co-carcinogenic effect with aflatoxins.[39,44,45] The tumor promoting effect of cyclopropenoid fatty acids may be dependent upon their incorporation into membranes.[46] Interestingly, malvalic and sterculic acids inhibit the growth of seed-eating lepidopteran larvae and may be part of the defense of cotton plants against these insects.[47] These fatty acids may also be effective antifungal agents, inhibiting the growth of some plant pathogenic fungi at concentrations that appear biologically relevant.[38,48] The most intensely studied of the

unusual fatty acids from a dietary viewpoint is erucic acid, due to fears that the consumption of rapeseed oil may be detrimental to human health. Chronic feeding of erucic acid to experimental animals has a range of deleterious effects,[49] but whether these are sufficiently severe to propose a herbivore-defense role for erucic acid in seeds is questionable.

It is possible that the use of unusual fatty acids as a carbon source may require adaptations of lipases or β-oxidation enzymes not present in the herbivore so that it cannot catabolise them and remains unrewarded for eating seeds in which unusual fatty acids make up a large component of stored carbon. The purgative properties of castor oil, in which triricinolein is the predominant lipid, are well known. However the significance of ricinoleic acid in this context is unclear since the castor seed seems already well protected by the presence of toxic and allergenic proteins.

The question of why so many unusual fatty acids (Table 1) have evolved in plants may be considered as a subset of the same question concerning the extreme diversity of plant secondary metabolites.[3,50] Indeed, in some cases, unusual fatty acids may be starting points for plant secondary metabolism. For example, crepenynic acid (Table 2) is believed to be the precursor of most of the large range of polyacetylenes synthesized by a small group of plant families. These polyacetylenes may be involved in plant-plant or plant-animal interactions, and some have toxic or antibiotic properties. However, the presence of acetylenic fatty acids in other plant families such as the Santalaceae seems not to be accompanied by the accumulation of polyacetylenes.[15]

It is apparent that the accumulation of unusual fatty acids in lipid bodies of the seed does not incur any selective disadvantage. By contrast, some unusual fatty acids might be expected to disrupt membrane structure and function if incorporated into membrane-forming lipids. There is relatively little direct biological evidence for a disruptive effect of unusual fatty acids upon membrane structure. The few available studies have exploited mutants of *Escherichia coli* and of yeast which are incapable of fatty acid desaturation and require exogenous unsaturated fatty acids for growth. The fatty acids supplemented to the growth medium become incorporated in the membranes, and thus provide a technique for correlating fatty acid structure with ability to augment saturated fatty acids in the formation of functional membranes.

Data from such experiments are somewhat contradictory, perhaps due to toxic impurities in some of the fatty acids used. In yeast,[51-54] it appeared that *cis*-polyunsaturates supported growth regardless of double bond positions. *Cis*-monounsaturates appeared to be effective only for certain desaturation positions, though this was due to an incompatibility with long stretches of contiguous saturation rather than an incompatibility with certain bond positions. *Trans*-unsaturation was less effective than *cis*. Certain hydroxy fatty acids (e.g. ricinoleic) were compatible with growth, though they were less effective than the non-oxygenated analogues (e.g. oleic acid). The use of some unusual fatty acids in membranes was associated with their modification, such as the acetylation of hydroxystearic acids.[55] Results of studies with *E. coli* are broadly similar; branched chain, brominated and *trans*-unsaturated fatty acids could support growth, as could cyclopropanoid acids which are

in fact "usual" fatty acids for this organism[56-58] (see below). These studies show that at least some unusual fatty acids can exist in some functional membranes, but it remains quite possible that their inclusion in normal plant membranes would be deleterious.

A new technique for modifying plant membrane fatty acid composition *in vivo* has been described,[59] which might be used to address the question of the compatibility of unusual fatty acids with membrane function. When fatty acids (15:0, 17:0, 17:1, 18:1) were applied as their Tween esters to leaves or other organs they were extensively incorporated into membrane lipids. Techniques also exist for the study of physical properties of multilamellar liposomes containing unusual fatty acids, generated *in vitro*.[60,61] One such study[60] showed that cyclopropanoid fatty acids are in fact eminently suited to the formation of functional membranes stable over a broad temperature range. As mentioned above, such fatty acids are normal components of some bacterial membranes.[62,63]

GENERAL BIOCHEMISTRY OF UNUSUAL FATTY ACID ACCUMULATION IN PLANTS

In addition to the presence of enzymes involved directly in the synthesis of unusual fatty acids, plants which accumulate unusual fatty acids may require other specialized proteins. Germination of the seeds in which they occur requires that the catabolic enzymes, such as lipases and the enzymes of β-oxidation, must be able to

accept the unusual fatty acids and that unusual structures formed during β-oxidation can be processed. I am aware of only one pertinent study. When [^{14}C]ricinoleate was catabolised by homogenates of germinating pea and castor seeds, β-oxidation was blocked at the C_{10} level in pea, but went to completion in castor.[64] The C_{10} product identified in pea, 4-keto-decanoic acid, was presumably not further metabolizable, causing the arrest of β-oxidation at this point. A pathway was proposed for the degradation in castor of 4-hydroxy-decanoic acid via 2-hydroxy-octanoic acid or 4-keto-decanoic acid, 2-keto-octanoic acid, and heptanoic acid, but the operation of this pathway has not been verified.

As mentioned above, unusual fatty acids accumulate almost exclusively in the triacylglycerol fraction, and are in some way excluded from the polar lipids. This is particularly intriguing since diacylglycerol is a precursor of both triacylglycerol and polar lipid. With castor microsomes, there was some indication that the pool of ricinoleoyl-containing polar lipid is minimised by a preference of diacylglycerol acyltransferase for ricinoleate-containing diacylglycerols.[65] A similar result was obtained with *Cuphea lanceolata* microsomes[66] where the diacylglycerol acyltransferase is highly active and selective for diacylglycerol containing medium-chain fatty acids. In addition, the lysophosphatidic acid acyltransferase was selective for both donor and acceptor acyl groups such that didecanoyl and dioleoyl species of phosphatidic acid accounted for the majority synthesized. The lysophosphatidic acid acyltransferase of palm (*Syagrus cocoides*) microsomes also preferentially acylates lysophosphatidic acid containing medium-chain (12:0) acyl groups with medium-chain

acyl-CoAs, again favouring dilauroylglycerol over mixed (12:0, 18:1) diacylglycerol.[67] In borage (*Borago officinalis*), γ-linolenic acid may be efficiently acylated to the *sn1*- and *sn2*- positions of glycerol-3-phosphate, but in fact accumulates preferentially at the *sn3*-position of triacylglycerol, and is altogether absent from the *sn-1* position. The diacylglycerol acyltransferase preferably uses γ-18:3-CoA and this may minimise the pool size of γ-18:3-CoA, such that γ-18:3 is concentrated in triacylglycerol.[68] Data accruing from a number of other studies[69-71] tend to strengthen the view obtained from those described above, that targetting of unusual fatty acids to triacylglycerol can be at least partly explained by the relative activities and selectivities of the acyltransferases.

In some cases the discrimination of acyltransferases actually limits the unusual fatty acid content of triacylglycerol. A well-known example is the "66% barrier" to erucate content in triacylglycerol of *Brassica* seeds. In this case the lysophosphatidic acid acyltransferase does not accept erucoyl-CoA as an acyl donor, limiting erucate to the *sn1*- and *sn3*-positions of triacylglycerol. This specificity is observed in several species but the enzyme from meadowfoam (*Limnanthes alba*) has been recognised as an exception.[72] An understanding of enzyme specificities and the ability to exploit those enzymes with desirable properties, such as the meadowfoam lysophosphatidic acid acyltransferase, will be important facets in the future manipulation of oil crops through molecular techniques. Likewise, it is important to investigate the metabolic fate of unusual fatty acids introduced (*in vitro*, and eventually through molecular techniques) into plants lacking unusual fatty acids. One such study[73] showed that the

targetting of unusual fatty acids to triacylglycerol is maintained, but the rate of esterification of the alien fatty acids was slow.

CASTOR (*RICINUS COMMUNIS*) AS A SOURCE OF RICINOLEIC ACID

Castor (*Ricinus communis* L.) is a minor oilseed crop, the large seeds being harvested for their oil content (~50%) which is rich (85-90% of total fatty acids) in the hydroxylated fatty acid, ricinoleic acid (12D-hydroxyoctadec-*cis*-9-enoic acid).[74]

Castor is a widely variable plant of warmer climates, and is the only species of the genus *Ricinus*, in the Euphorbiaceae (spurge family) (for classifications of this family, see references 75-77). Castor probably originated in the Ethiopian-East African region, and has been cultivated for millenia, although it remains semi-wild in many regions, from which much of the total production of developing countries may be harvested.[78] Castor has 2n=20 chromosomes, with some claims that it is an allopolyploid (x=5). The plants are generally monoecious, bearing racemes separated into upper female flowers and lower male flowers. Castor is somewhat unusual among dicots in that the endosperm, rather than the cotyledons, is the storage tissue.

Oil pressed or extracted from castor seeds has hundreds of industrial uses, many based upon the properties endowed by the hydroxylated fatty acid. The most important uses are production of: paints and varnishes, nylon-type synthetic polymers, resins, lubricants, and cosmetics.[78] In addition to oil, the castor seed contains the extremely toxic protein ricin, allergenic proteins,[79] and the pyridine alkaloid ricinine.[80]

These constituents preclude (see reference 81) the use of the seed meal (following oil extraction) as a livestock feed, normally an important economic aspect of oilseed utilisation. Furthermore, with the variable nature of castor plants and a lack of investment in breeding, castor has few favourable agronomic characteristics. For a combination of these reasons, castor is no longer grown in the United States. In 1965, US production of castor oil was approximately 65 million pounds, from around 80 000 acres cropped. Current annual imports of about 92 million pounds cost around $45 million.[82]

The major producers of castor oil are India, Brazil and China. Supplies have been somewhat unstable and subject to artificial pricing, and regain of a domestic source of castor oil would be attractive.[82] The production of ricinoleic acid, the important constituent of castor oil, in an established oilseed crop (such as sunflower or rapeseed) through genetic engineering would be a particularly effective means of circumventing the problems of castor production.

RESEARCH GOAL

The difference between ricinoleic acid and the usual plant fatty acid, linoleic acid, is simple, the acyl chain being hydroxylated at the C12 position instead of the usual desaturation at this position. The goal of this work was to isolate a gene or genes necessary for this hydroxylation in castor, with a view toward eventual production of the hydroxylated fatty acid in a different plant.

REFERENCES

1. Hitchcock, C., and Nichols, B.W., *Plant Lipid Biochemistry*, Academic Press, London, 1971, chap. 1.

2. Gunstone, F.D., Harwood, J.L., and Padley, F.B., *The Lipid Handbook*, Chapman and Hall, London, 1986, chaps. 1 & 3.Webster, G.L., The saga of the spurges: a review of classification and relationships in the Euphorbiales, *Bot. J. Linn. Soc.*, 94, 3-46, 1987.

3. Wink, M., Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores, *Theor. Appl. Genet.*, 75, 225, 1988.

4. Kleiman, R., and Payne-Wahl, K.L., Fatty acid composition of seed oils of the Meliaceae, including one genus rich in cis-vaccenic acid, *J. Am. Oil Chem. Soc.*, 61, 1836, 1984.

5. Shibahara, A., Yamamoto, K., Takeoka, M., Kinoshita, A., Kajimoto, G., Nakayama, T., and Noda, M., Application of a GC-MS method using deuterated fatty acids for tracing cis-vaccenic acid biosynthesis in kaki pulp, *Lipids*, 24, 488, 1989.

6. Radunz, A., On the function of methyl-branched chain fatty acids in phospholipids of cell membranes in higher plants, in *The Metabolism, Structure, and Function of Plant Lipids*, Stumpf, P.K., Mudd, J.B., and Nes, W.D., Eds., Plenum, New York, 1987, 197.

7. Hitchcock, C., Structure and distribution of plant acyl lipids, in *Recent Advances in the Chemistry and Biochemistry of Plant Lipids*, Galliard, T. and Mercer, E.I., Eds., Academic Press, London, 1975, chap. 1.

8. Smith, C.R., Jr., Occurrence of unusual fatty acids in plants, in *Progress in the Chemistry of Fats and Other Lipids*, Vol. 11, pt 1, Holman, R.T., Ed., 1970.

9. Markley, K.S., *Fatty Acids; Their Chemistry, Properties, Production, and Uses*, Vol. 1, 2nd ed., Krieger, Malabar, 1983.

10. Smith C.R., Jr., Unusual seed oils and their fatty acids, in *Fatty Acids*, Pryde E.H., Ed., American Oil Chemists' Society, Champaign, 1979, chap. 2.

11. Hopkins, C.Y., Fatty acids with conjugated unsaturation, in *Topics in Lipid Chemistry*, Vol. 3, Gunstone, F.D., Ed., John Wiley & Sons, New York, 1972, chap. 2.

12. Browse, J., and Somerville, C., Glycerolipid synthesis: biochemistry and regulation, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 42, 467, 1991.

13. Spener F., and Tober, I., Cyclopentenylfettsauren im Brennpunkt, *Fette, Seifen, Anstrichm.*, 83, 401, 1981.

14. Tober I., and Conn, E.E., Cyclopentenylglycine, a precursor of deidaclin in Turnera ulmifolia, *Phytochemistry*, 24, 1215, 1985.

15. Harborne, J.B., and Turner, B.L., *Plant Chemosystematics*, Academic Press, London, 1984, chap. 8.

16. Brieskorn, C.H., and Kabelitz, L., Hydroxyfettsauren aus dem cutin des blattes von Rosmarinus officinalis, *Phytochemistry*, 10, 3195, 1971.

17. Green, A.G., The occurrence of ricinoleic acid in Linum seed oils, *J. Am. Oil Chem. Soc.*, 61, 939, 1984.

18. Garg, S.P., Sherwani, M.R.K., Arora, A., Agarwal, R., and Ahmad, M., Ricinoleic acid in Vinca rosea seed oil, *J. Oil Technol. Assoc. India*, 19, 63, 1987.

19. Ahmad, M.U., Husain, S.K., and Osman, S.M., Ricinoleic acid in Phyllanthus niruri seed oil, *J. Am. Oil Chem. Soc.*, 58, 673, 1981.

20. Daulatabad, C.D., and Mirajkar, A.M., Ricinoleic acid in Artocarpus integrifolia seed oil, *J. Am. Oil Chem. Soc.*, 66, 1631, 1989.

21. Daulatabad, C.D., Desai, V.A., Hosamani, K.M., and Jamkhandi, A.M., Novel fatty acids in Azima tetracantha seed oil, *J. Am. Oil Chem. Soc.*, 68, 978, 1991.

22. Daulatabad, C.D. and Hosamani, K.M.; Unusual fatty acids in Brunfelsia americana seed oil: a rich source of oil, *J. Am. Oil Chem. Soc.*, 68, 608, 1991.

23. Yunusova, S.G., Gusakova, S.D., Glushenkova, A.I., Lipids from the seeds of Securinega suffruticosa, *Khim. Prir. Soedin. (Tashk.)*, 3, 277, 1986.

24. Kleiman, R., Spencer, G.F., Earle, F.R., Nieschlag, H.J., and Barclay, A.S., Tetra-acid triglycerides containing a new hydroxy eicosadienoyl moiety in Lesquerella auriculata seed oil, *Lipids*, 7, 660, 1972.

25. Sreenivasan, B., Kamath, N.R., and Kane, J.G., Studies on castor oil. 1. Fatty acid composition of castor oil, *J. Am. Oil Chem. Soc.*, 33, 61, 1956.

26. Badami, R.C., and Kudari, S.M., Analysis of Hiptage madablota seed oil, *J. Sci. Food Agric.*, 21, 248, 1970.

27. Nikolin, A., Nikolin, B., and Jankovic, M., Ipopurpuroside, a new glycoside from Ipomoea purpurea, *Phytochemistry*, 17, 451, 1978.

28. Rao, S.V., Paulose, M.M., and Vijayalakshmi, B., A note on the fatty acids present in oilseed phospholipids, *Lipids*, 2, 88, 1967.

29. Persmark, U., Main constituents of rapeseed lecithin, *J. Am. Oil Chem. Soc.*, 45, 742, 1968.

30. Vijayalakshmi, B., and Rao, S.V., Fatty acid composition of phospholipids in seed oils containing unusual acids, *Chem. Phys. Lipids*, 9, 82, 1972.

31. Kondoh, H., and Kawabe, S., Phospholipids in castor seeds, *Agr. Biol. Chem.*, 39, 745, 1975.

32. Nasirullah, Werner, G., and Seher, A., Fatty acid composition of lipids from edible parts and seeds of vegetables, *Fette, Seifen, Anstrichm.*, 86, 264, 1984.

33. Prasad, R.B.N., Rao, Y.N., and Rao, S.V., Phoshpholipids of palash (Butea monosperma), papaya (Carica papaya), jangli badam (Sterculia foetida), coriander (Coriandrum sativum) and carrot (Daucus carota) seeds, *J. Am. Oil Chem. Soc.*, 64, 1424, 1987.

34. Jamieson, G.R., and Reid, E.H., Analysis of oils and fats by gas chromatography. V. Fatty acid composition of the leaf lipids of Myosotis scorpioides, *J. Sci. Food Agric.*, 19, 628, 1968.

35. Jamieson, G.R., and Reid, E.H., The leaf lipids of some members of the Boraginaceae family, *Phytochemistry*, 8, 1489, 1969.

36. Appelqvist, L.-A., Biochemical and structural aspects of storage and membrane lipids in developing oilseeds, in *Recent Advances in the Chemistry and Biochemistry of Plant Lipids*, Galliard, T. and Mercer, E.I., Eds., Academic Press, London, 1975, chap. 8.

37. Shenstone, F.S., and Vickery, J.R., Occurrence of cyclo-propene fatty acids in some plants of the order Malvales, *Nature*, 190, 168, 1961.

38. Schmid, K.M., and Patterson, G.W., Distribution of cylclopropenoid fatty acids in malvaceous plant parts, *Phytochemistry*, 27, 2831, 1988.

39. Mangold, H.K., and Spener, F., Biosynthesis of cyclic fatty acids, in *The Biochemistry of Plants*, Vol. 4, Stumpf, P.K., Ed., Academic Press, London, 1980, chap. 19.

40. Lie Ken Jie, M.S.F., and Sinha, S., Fatty acid composition and the characterization of a novel dioxo C_{18}-fatty acid in the latex of Hevea brasiliensis, *Phytochemistry*, 20, 1863, 1981.

41. Gandhi, V.M., Cherian, K.M., and Mulky, M.J., Nutritional and toxicological evaluation of rubber seed oil, *J. Am. Oil Chem. Soc.*, 67, 883, 1990.

42. Harper, D.B., and Hamilton, J.T.G., Identification of threo-18-fluoro-9,10-dihydroxystearic acid: a novel ω-fluorinated fatty acid from Dichapetalum toxicarium seeds, *Tetrahedron Lett.*, 31, 7661, 1990.

43. Abdel-Moety, E.M., Cyclopentenylfettsauren als ausgangsmaterial zur gewinnung neuer wirkstoffe, *Fette, Seifen, Anstrichm.*, 83, 65, 1981.

44. Greenberg, A., and Harris, J., Cyclopropenoid fatty acids, *J. Chem. Edu.*, 59, 539, 1982.

45. Schuch, R., and Ahmad, F., Structure and biological significance of triacylglycerols containing cyclopropene acyl moieties, *Fat Sci. Technol.*, 89, 338, 1987.

46. Pawlowski, N.E., Hendricks, J.D., Bailey, M.L., Nixon, J.E., and Bailey G.S., Structural-bioactivity relationship for tumor promotion by cyclopropenes, *J. Agric. Food Chem.*, 33, 767, 1985.

47. Chan, B.G., Waiss, A.C., Jr., Binder, R.G., and Elliger, C.A., Inhibition of lepidopterous larval growth by cotton constituents, *Entomol. Exp. Appl.*, 24, 294, 1978.

48. Schmid, K.M., and Patterson, G.W., Effects of cyclopropenoid fatty acids on fungal growth and lipid composition, *Lipids*, 23, 248, 1988.

49. Kramer, J.K.G., Sauer, F.D., and Pigden, W.J., *High and Low Erucic Acid Rapeseed Oils*, Academic Press, Toronto, 1983, chaps. 11-21.

50. Williams, D.H., Stone, M.J., Hauck, P.R., and Rahman, S.K., Why are secondary metabolites (natural products) biosynthesized?, *J. Nat. Prod.*, 52, 1189, 1989.

51. Proudlock, J.W., Haslam, J.M., and Linnane, A.W., Biogenesis of mitochondria 19. The effects of unsaturated fatty acid depletion on the lipid composition and energy metabolism of a fatty acid desaturase mutant of Saccharomyces cerevisiae, *Bioenergetics*, 2, 327, 1971.

52. Wisnieski, B.J., and Kiyomoto, R.K., Fatty acid desaturase mutants of yeast: growth requirements and electron spin resonance spin-label distribution, *J. Bacteriol.*, 109, 186, 1972.

53. Lands, W.E.M., Sacks, R.W., Sauter, J., and Gunstone, F., Selective effects of fatty acids upon cell growth and metabolic regulation, *Lipids*, 13, 878, 1978.

54. Nes, W.D., Adler, J.H., and Nes, W.R., A structure-function correlation for fatty acids in Saccharomyces cerevisiae, *Exp. Mycol.*, 8, 55, 1984.

55. Light, R.J., Lennarz, W.J., and Bloch, K., The metabolism of hydroxystearic acids in yeast, *J. Biol. Chem.*, 237, 1793, 1962.

56. Silbert, D.F., Ruch, R., and Vagelos, P.R., Fatty acid replacements in a fatty acid auxotroph of Escherichia coli, *J. Bacteriol.*, 95, 1658, 1968.

57. Silbert, D.F., Ladenson, R.C., and Honegger, J.L., The unsaturated fatty acid requirement in Escherichia coli. Temperature dependence and total replacement by branched-chain fatty acids, *Biochim. Biophys. Acta*, 311, 349, 1973.

58. Machtiger, N.A., and Fox, C.F., Biochemistry of bacterial membranes, *Ann. Rev. Biochem.*, 42, 575, 1973.

59. Terzaghi, W.B., Manipulating membrane fatty acid compositions of whole plants with tween-fatty acid esters, *Plant Physiol.*, 91, 203,

1989.

60. Dufourc, E.J., Smith, I.C.P., and Jarrell, H.C., Role of cyclopropane moieties in the lipid properties of biological membranes: a ^2H NMR structural and dynamical approach, *Biochemistry*, 23, 2300, 1984.

61. Isaacson, Y., Riehl, T.E., and Stenson, W.F., Nonelectrolyte permeability of liposomes of hydroxyfatty acid-containing phosphatidylcholines, *Biochim. Biophys. Acta*, 986, 295, 1989.

62. Grogan, D.W., and Cronan, J.E., Cloning and manipulation of the Escherichia coli cyclopropane fatty acid synthase gene: physiological aspects of enzyme overproduction, *J. Bacteriol.*, 158, 286, 1984.

63. Grogan, D.W., and Cronan, J.E., Characterization of Escherichia coli mutants completely defective in synthesis of cyclopropane fatty acids, *J. Bacteriol.*, 166, 872, 1986.

64. Hutton, D., and Stumpf, P.K., Fat metabolism in higher plants LXII. The pathway of ricinoleic acid catabolism in the germinating castor bean (Ricinus communis L.) and pea (Pisum sativum L.), *Arch. Biochem. Biophys.*, 142, 48, 1971.

65. Bafor, M., Smith, M.A., Jonsson, L., Stobart, K., and Stymne, S., Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (Ricinus communis) endosperm, *Biochem. J.*, 280, 507, 1991.

66. Bafor, M., Jonsson, L., Stobart, A.K., and Stymne, S., Regulation of triacylglycerol biosynthesis in embryos and microsomal preparations from the developing seeds of Cuphea lanceolata, *Biochem. J.*, 272, 31, 1990.

67. Oo, K.-C., and Huang, A.H.C., Lysophosphatidate acyltransferase activities in the microsomes from palm endosperm, maize scutellum, and rapeseed cotyledon of maturing seeds, *Plant Physiol.*, 91, 1288, 1989.

68. Griffiths, G., Stobart, A.K., and Stymne, S., Δ^6- and Δ^{12}-desaturase activities and phosphatidic acid formation in microsomal preparations from the developing cotyledons of common borage (Borago officinalis), *Biochem. J.*, 252, 641, 1988.

69. Mukherjee, K.D., and Kiewitt, I., Changes in fatty acid composition of

lipid classes in developing mustard seed, *Phytochemistry*, 23, 349, 1984.

70. Cao, Y.-Z., and Huang, A.H.C., Acyl coenzyme A preference of diacylglycerol acyltransferase from the maturing seeds of Cuphea, maize, rapeseed, and canola, *Plant Physiol.*, 84, 762, 1987.

71. Fehling, E., and Mukherjee, K.D., Biosynthesis of triacylglycerols containing very long chain mono-unsaturated fatty acids in seeds of Lunaria annua, *Phytochemistry*, 29, 1525, 1990.

72. Cao, Y.-Z., Oo, K.-C., and Huang, A.H.C., Lysophosphatidate acyltransferase in the microsomes from maturing seeds of meadowfoam (Limnanthes alba), *Plant Physiol.*, 94, 1199, 1990.

73. Battey, J.F., and Ohlrogge, J.B., A comparison of the metabolic fate of fatty acids of different chain lengths in developing oilseeds, *Plant Physiol.*, 90, 835, 1989.

74. Weiss, E.A., *Oilseed Crops,* Longman, London, 1983.

75. Pax, F., and Hoffmann, K., Euphorbiaceae, in *Die Naturlichen Pflanzenfamilien, Zweite Auflage, 19c,* Engler, A., and Prantl, K., Eds., W. Engelmann, Leipzig, 1931, 11-233.

76. Webster, G.L., Conspectus of a new classification of the Euphorbiaceae, *Taxon,* 24, 593-601, 1975.

77. Webster, G.L., The saga of the spurges: areview of classification and relationships in the Euphorbiales, *Bot. J. Linn. Soc.*, 94, 3-46, 1987.

78. Atsmon, D., Castor, in *Oil Crops of the World,* Robbelen, G., Downey, K.R., and Ashri, A., Eds., McGraw-Hill, New York, 1989, chap. 24.

79. Kanerva, L., Estlander, T., and Jolanki, R., Long-lasting contact urticaria from castor bean, *J. Am. Acad. Dermatology,* 33, 351-355, 1990.

80. Rizk, A.-F.M., The chemical constituents and economic plants of the Euphorbiaceae, *Bot. J. Linn. Soc.*, 94, 293-326, 1987.

81. Horton, J., and Williams, M.A., A cooker-extruder for deallergenation of castor bean meal, *J. Am. Oil Chem. Soc.*, 66, 227-231, 1989.

82. Hawkins, R., Bring back the U.S. castor crop!, *J. Am. Oil Chem. Soc.*, 66,

1562-1563, 1989.

CHAPTER 2

BIOCHEMICAL STUDIES

ABSTRACT

Experiments aimed at a better understanding of oleate-12-hydroxylase are described. An earlier report[1] suggested that hydroxylase activity arises in the castor seed at a stage in development subsequent to the differentiation of the cellular endosperm and the onset of lipid deposition. Results are presented contradicting this report: hydroxylase activity was readily detectable at all stages of seed development in which cellular endosperm was present.

Inhibition of oleate-12-hydroxylase activity following incubation with anti-cytochrome b_5 immunogobulins demonstrated that cytochrome b_5 is the electron donor to the hydroxylase.

Oleate-12-hydroxylase activity was lost during various attempts to fractionate active endosperm extracts, thwarting efforts at partial purification.

In an alternative strategy, rabbits were immunised with denatured endosperm proteins. Polyclonal antibodies produced, if inhibitory to oleate-12-hydroxylase, could be purified and used to isolate the hydroxylase protein. However, these antibodies had no effect on hydroxylase activity.

Biochemical features of the hydroxylase are summarised and compared to features of the microsomal fatty acid desaturases.

INTRODUCTION

The biosynthesis of ricinoleic (12D-hydroxyoctadec-*cis*-9-enoic) acid from oleic acid in the developing endosperm of castor (*Ricinus communis*) is relatively well studied. Morris[2] established in elegant double-labeling studies that hydroxylation occurs directly by hydroxyl substitution rather than via an unsaturated-, keto- or epoxy-intermediate. Developing endosperm slices were incubated with a mixture of [1-^{14}C]oleic acid and *erythro*-12,13-ditritio-oleic acid. The ^{3}H/^{14}C ratio of ricinoleate synthesized was found to be 75% of that of the substrate mixture, and upon chemical oxidation to 12-keto-oleate the ratio was 55%, close to a predicted 50%. These results can only be obtained by a hydroxyl substitution mechanism (^{3}H/^{14}C ratios in the ricinoleate and 12-keto-oleate would be, respectively, for a keto-intermediate: 50% and 50%, for an unsaturated intermediate: 50% and 25%, and for an epoxy-intermediate also 50% and 25%).

Hydroxylation using oleoyl-CoA as precursor can be demonstrated in crude preparations or microsomes, but activity in microsomes is unstable and variable, and isolation of the microsomes involved a considerable, or sometimes complete loss of activity.[3,4] Oleic acid can replace oleoyl-CoA as a precursor, but only in the presence of CoA, Mg^{2+} and ATP,[3] which suggests that activation to the acyl-CoA is necessary. Furthermore, radioactivity from oleoyl-CoA is rapidly lost during hydroxylation assays, appearing first in lipid-esterified oleate, and subsequently in lipid-esterified ricinoleate. This and other experiments demonstrated a better substrate-product

relationship between lipid-esterified oleate and lipid-esterified ricinoleate than oleoyl-CoA and lipid-esterified ricinoleate, and no radioactivity could be detected in ricinoleoyl-CoA.[4] The hydroxylase is sensitive to cyanide and azide, and dialysis against metal chelators reduces activity, which could be restored by addition of $FeSO_4$, suggesting iron involvement in enzyme activity,[3] although contradictory results have also been reported.[4] There is no conversion of oleoyl-CoA to ricinoleic acid in the absence of molecular oxygen,[3,4] although earlier studies had indicated otherwise.[1,5] Hydroxylation requires NAD(P)H, and NADPH supports lower rates of hydroxylation than NADH.[3,4] Carbon monoxide does not inhibit hydroxylation, suggesting that a cytochrome P450 is not involved.[3,4] Data from a study of the substrate specificity of the hydroxylase show that all substrate parameters (i.e. chain length and double bond position with respect to both ends) are important; deviations in these parameters caused reduced activity relative to oleic acid.[6] The position at which the hydroxyl was introduced however, was determined by the position of the double bond, always being three carbons distal.

Fatty acid hydroxylation in organisms other than castor is now considered for comparison. The only other organism in which ricinoleic acid biosynthesis has been investigated is the ergot fungus, *Claviceps purpurea*. Ricinoleate accumulates (up to 40% of the fatty acids) in the glycerides produced particularly by sclerotia of anaerobic cultures.[7] As this suggests, oxygen is not necessary for the synthesis of ricinoleic acid in *Claviceps*, and the precursor of ricinoleic acid in fact appears to be linoleic acid.[8] However, ricinoleic acid may not be formed simply by hydration of

linoleic acid, since there are no free hydroxyl groups in ergot oil. Rather, the hydroxyl groups are all esterified to other, non-hydroxy fatty acids, leading to a range of tetra-acyl-, penta-acyl- and hexa-acyl-glycerides. These estolides may be formed by a direct enzymic addition of non-hydroxy fatty acids across the Δ12 double bond of linoleate.[9] Ricinoleic acid may, therefore, be merely an artifact of the hydrolysis employed to study the fatty acid composition of the oil.

Investigations of the biosynthesis of isoricinoleic acid (9-hydroxyoctadec-*cis*-12-enoic acid) in *Wrightia* spp.[10,11] have provided no conclusive evidence relating to the pathway of biosynthesis. Interestingly, 9-OH-18:0 is detected in the seed oil, and can be labelled after incubation of developing seed halves with [^{14}C]acetate, indicating that the substrate for hydroxylation may be stearate (by direct hydroxylation) or oleate (by double bond hydration). These workers[11] favoured the latter possibility because under their conditions [^{14}C]18:1 and [^{14}C]18:2 could act as precursors for [^{14}C]9-OH-18:1, and this conversion was greater under anaerobic conditions. However it seems that these anaerobic conditions may not have been rigorous, and incubation times were very long; indeed the conversion of [^{14}C]18:2 to [^{14}C]9-OH-18:0 and [^{14}C]9-OH-18:1 at equivalent rates must be viewed with some concern.

Little information pertinent to hydroxylation in castor is given by these comparisons with other organisms. The state of knowledge of the biochemical nature of the castor hydroxylase is summarised as follows. Oleate is the precursor of ricinoleic acid, and hydroxylation procedes by a direct hydroxyl susbstitution mechanism. The immediate enzyme substrate may be lipid-esterified oleate, and

some activity may be found in membrane fractions, but evidence that the enzyme is membrane-bound rather than soluble is weak. NAD(P)H is required, as probably is molecular oxygen. Preference for NADH over NADPH and insensitivity to carbon monoxide mitigate against the involvement of a cytochrome P450. No purification or partial purification of the enzmye has been reported. The present studies of hydroxylation in castor therefore commenced with the objective of acquiring a better knowledge of the biochemical nature of the hydroxylase, particularly focused on any information which may suggest methods of purification which could further the stated goal of cloning a gene(s) for the hydroxylase enzyme.

MATERIALS AND METHODS

Plant Material

Castor (*Ricinus communis* L. cv Baker 296) plants were grown throughout the year in the greenhouse. Developing inflorescences were removed for isolation of endosperm tissue in the laboratory. Cellular endosperm and embryo were removed from seeds at development stage III to stage V,[12] identified chiefly by the phase of rapid expansion of the opaque white endosperm. The tissues were isolated directly into liquid nitrogen, ground to a powder in a mortar and pestle, and could be stored at -80°C. This frozen powder is hereafter referred to as "endosperm".

Oleate-12-Hydroxylase Assays

Developmental Analysis of Ricinoleic Acid Biosynthesis. Frozen endosperm was ground with two volumes of extraction buffer (50 mM PIPES-KOH pH 7.1, 10% (w/v) sucrose, 1 mM cysteine, 5 μg ml^{-1} leupeptin) in a glass-in-glass homogenizer. The homogenate was micro-centrifuged ($\sim$ 13 000 g) for 15 min, and the supernatant was decanted from the pellet and fat pad and used for hydroxylase assays. Each assay contained enzyme (50 μl), assay buffer (50 mM PIPES-KOH pH 7.1, 10% (w/v) sucrose), NADH (0.4 mM) and 1-^{14}C-oleoyl-CoA (80 000 dpm, 52 Ci mol^{-1}), in a total volume of 0.5 ml. Assay tubes were incubated at 30°C for 60 min, then stopped by the addition of 15% methanolic KOH (0.5 ml). Lipids were saponified by heating to 80°C for 30 min, followed by recovery of the free fatty acids by neutralising with 2.5 M HCl (0.5 ml) and extraction with 3.5 ml hexane:isopropanol (3:2)/ 2.5 ml 0.2 M Na_2SO_4. The organic phase was dried under nitrogen and redissolved in 50 μl chloroform:methanol (2:1) for spotting (alongside authentic oleic acid and ricinoleic acid standards) on a silica TLC plate (Baker Si250). The plate was developed in a paper-lined tank containing benzene:ethyl ether:ethanol (100:30:2). Fatty acids were detected by staining with iodine vapour. Radioactivity could be quantitated by autoradiography followed by scraping and scintillation-counting of the oleic acid and ricinoleic acid spots, but equivalent quantitation was routinely obtained using a BioScan 2000 scanner.

Several lines of evidence demonstrate that the radioactive compound

quantitated in these assays is truly ricinoleic acid. When reaction products were esterified with methanol, and separated by TLC in a petroleum ether:ethyl ether (1:1) system they again co-migrated with authentic standards. Furthermore, in the standard solvent (C_6H_6:Et_2O:EtOH), a predominant iodine-staining spot (ricinoleic acid is a predominant fatty acid of the developing castor endosperm) readily overloads the plate, resulting in a tailing pattern that is precisely matched by tailing of the radioactivity thought to reside in ricinoleic acid. Finally, unequivocal proof of ricinoleic acid as the reaction product was obtained by using a tri-deuterated substrate in parallel to the radioactive substrate and analysis of the products by gas chromatography-mass spectrometry. This work was a collaboration with Dr. E.W. Underhill and is presented in the Appendix.

Electron Transfer to Oleate-12-Hydroxylase. Endosperm was extracted and assayed as described above, with the exception of the following changes. The extraction buffer contained only 50 mM PIPES-NaOH, pH 7.1, and centrifugation was at 1000g (SS-34 rotor) for 15 min. The supernatant was desalted by collecting turbid fractions from a small Sephadex G-25M column equilibrated with the extraction buffer. The same enzyme preparation was stored without loss of activity by freezing drop-wise in liquid nitrogen and maintained at -80°C, and used for this series of experiments. The assay buffer was neutralised with NaOH in place of KOH, and contained 1 mg ml^{-1} fatty acid free bovine serume albumin in addition to the other components.

Polyclonal Antibodies Raised Against Crude Enzyme Extracts. Endosperm was extracted and assayed as described above, with the exception of the following changes. The extraction buffer contained only 50 mM PIPES-NaOH, pH 7.1, and leupeptin (5 μg ml^{-1}), and centrifugation was at 1000g (SS-34 rotor) for 10 min. Assays used only 25 μl enzyme, and contained no sucrose.

Protein Determination

Protein was assayed by the method of Bradford,[13] using BioRad reagents, with bovine serum albumin as standard.

Polyclonal Antibodies raised against crude enzyme extracts

Endosperm was extracted as described above, but the supernatant was further processed as follows. $MgCl_2$ (1M) was added to a final concentration of 50 mM, and the preparation was micro-centrifuged for 10 min. The supernatant was passed through an 0.2 μm filter, and the filtrate desalted by chromatography on Sephadex-G25M as described above. This preparation was active in hydroxylase assays, and was loaded (3.7 mg protein per gel) onto two preparative (1.5 mm) polyacrylamide (10%) gels.[14] Following electrophoresis, protein was transferred[15] to nitrocellulose filters. Strips of the filters were baked at 80°C *in vacuo* for 30 min, chopped finely with a razor, followed by homogenization in a glass-in-glass homogenizer with

deionised water. The nitrocellulose powder was pelleted and resuspended in sterile phosphate-buffered saline solution.[16] These samples were divided into three aliquots, and used to immunise female New Zealand White rabbits. The rabbits were bled one day prior to injection, injected subcutaneously,[17] then boosted one month later, and again after a further three weeks. Immune sera used in the experiments described were obtained one week after the second boost, and had higher titre (as determined by signal strength on western blots) than sera obtained one week after the first boost.

Serum was cleared at 5000g for 10 min and purified by ion exchange chromatography with DEAE-Sephacel as described[18] and stored at -20°C. Further purification by affinity chromatography on protein G-Separose[19] was necessary prior to incubations with enzyme, to remove presumed proteases.

A western blot[15] was prepared from a preparative (0.75 mm) polyacrylamide (10%) gel containing protein (0.4 mg total) from the same enzyme preparation as used to immunise the rabbits. The blot was cut into strips (each representing 40 μg protein loaded) and blocked with BSA. Each strip was then incubated with diluted (1/200) immune serum or preimmune serum (from the same rabbit), and developed using goat-anti-rabbit alkaline phosphatase (Kirkegaard & Perry) with NBT/BCIP.[16]

Staphylococcus aureus cells were the gift of Dr. John Shanklin, prepared according to Kessler[20], and were washed extensively with assay buffer prior to use.

DEVELOPMENTAL ANALYSIS OF RICINOLEIC ACID BIOSYNTHESIS

A report in the literature[1] suggests that oleate-12-hydroxylase becomes active in the developing castor endosperm at a stage subsequent to and distinct from the differentiation of the storage endosperm and the commencement of storage lipid synthesis. This could be exploited by examining the possibly small number of proteins which appear in the endosperm concomitant with the onset of oleate hydroxylation. An initial experiment was therefore done to confirm the previous observations.[1]

A comprehensive descriptive morphology of the developing castor seed has been developed by Greenwood and Bewley.[12] In this timetable the cellular endosperm, which is the major lipid accumulating tissue, first arises at stage III (heart shaped embryo). Lipid accumulates very rapidly from stages IV (early cotyledon) to VII (full cotyledon),[21] concomitant with the rapid enlargement of the cellular endosperm. Lipid accumulates more gradually during stages VII to X (maturation).

Endosperms were dissected from developing inflorescences removed from greenhouse-grown plants. Inner integument, nucellus and free nuclear endosperm were isolated from stage II seeds. Cellular endosperm was isolated from stage III seeds, with a minimum of adhering nucellus and inner integument. Stage IV to stage VIII cellular endosperms were isolated together with the enclosed embryo. The isolated tissues were frozen, homogenised and assayed for hydroxylase activity and protein content.

The conversion of C^{14}-oleoyl-CoA to C^{14}-ricinoleate is shown in Table 3 for each of the developmental stages. No hydroxylase activity was detectable in tissues from stage II, which precedes the differentiation of the cellular endosperm in which ricinoleate-containing lipid accumulates. However, hydroxylase activity was measured in cellular endosperm from stage III and all subsequent stages, demonstrating that ricinoleic acid biosynthetic capacity arises in parallel with the storage tissue.

This result demonstrates that there is no period of lipid accumulation prior to the onset of oleate hydroxylation, contrary to the observations of James *et al.*[1] Rather, all stages of the seed in which the cellular endosperm is present and triglyceride synthesis occurs are also competent for oleate hydroxylation. Since the appearance of the oleate hydroxylase is concomitant with the differentiation of the cellular endoperm and the host of biosynthetic capacities which characterise it, it was decided not to pursue an examination of protein profiles in the seed before and after the appearance of the oleate hydroxylase.

ELECTRON TRANSFER TO OLEATE-12-HYDROXYLASE

Castor oleate-12-hydroxylase is dependent upon NAD(P)H.[3,4] This was also observed in the current study. When enzyme preparations were desalted by chromatography on Sephadex G-25M, activity could only be recovered by addition of NADH, or less effectively, NADPH. Another class of plant fatty acid oxidative

Table 3. Oleate-12-hydroxylase activity measured in extracts of endosperm isolated from castor seeds at different stages of development.

	Oleate-12-Hydroxylase Activity	
Stage[a]	**pmol ricinoleate/mg fw/h**	**pmol ricinoleate/mg protein/h**
II	0	0
III	0.14	18.2
IV	0.59	39.8
V	0.71	38.7
VI	0.62	48.3
VII	0.52	35.8
VIII	>0.38[b]	33.3

[a]Stages according to Greenwood and Bewley.[12]

[b]This value is a minimum bound to the true value, since a little of the enzyme preparation was spilled, precluding accurate calculation.

enzymes, the microsomal desaturases, also require NAD(P)H, and it has been shown[19] that donation of electrons originating from NADH to the desaturase is via cytochrome b_5. This is also the case for mammalian desaturases, where NADH-cytochrome b_5 reductase and cytochrome b_5 are involved in transfer of electrons to the desaturase.[22,23] Cytochromes P450 on the other hand, including plant fatty acid hydroxylases and epoxidases involved in cutin biosynthesis[24-30] typically utilize NADPH preferentially over NADH, and NADPH-cytochrome P450 reductase has been shown to be the electron donor to these enzymes.

An experiment was done to test whether cytochrome b_5 is also the electron donor to oleate-12-hydroxylase. This was by the same approach as Kearns *et al.*,[19] who used antibodies raised against cytochrome b_5 to hinder its interactions with either the NADH-cytochrome b_5 reductase, the desaturase, or both. Strong inhibition of oleate-12-hydroxylase by anti-cytochrome b_5 antibodies would be further evidence that this enzyme is not a cytochrome P450.

Preliminary experiments showed that maximal hydroxylase activity (20 pmol ricinoleate formed h^{-1}) was achieved with 111-125 μg of an enzyme preparation, more enzyme (to 277 μg) resulting in the same measured activity (Figure 1), indicating that these enzyme quantities were saturating. Furthermore, the enzyme could be stored for at least 5 h on ice before initiation of the assay without loss of activity (Figure 2).

Enzyme (83 μg protein) was incubated on ice for 6 h before assay with pre- or post-immune IgG from a mouse injected with purified cauliflower cytochrome b_5.[19] The results presented in Figure 3 demonstrate that while incubation with pre-immune

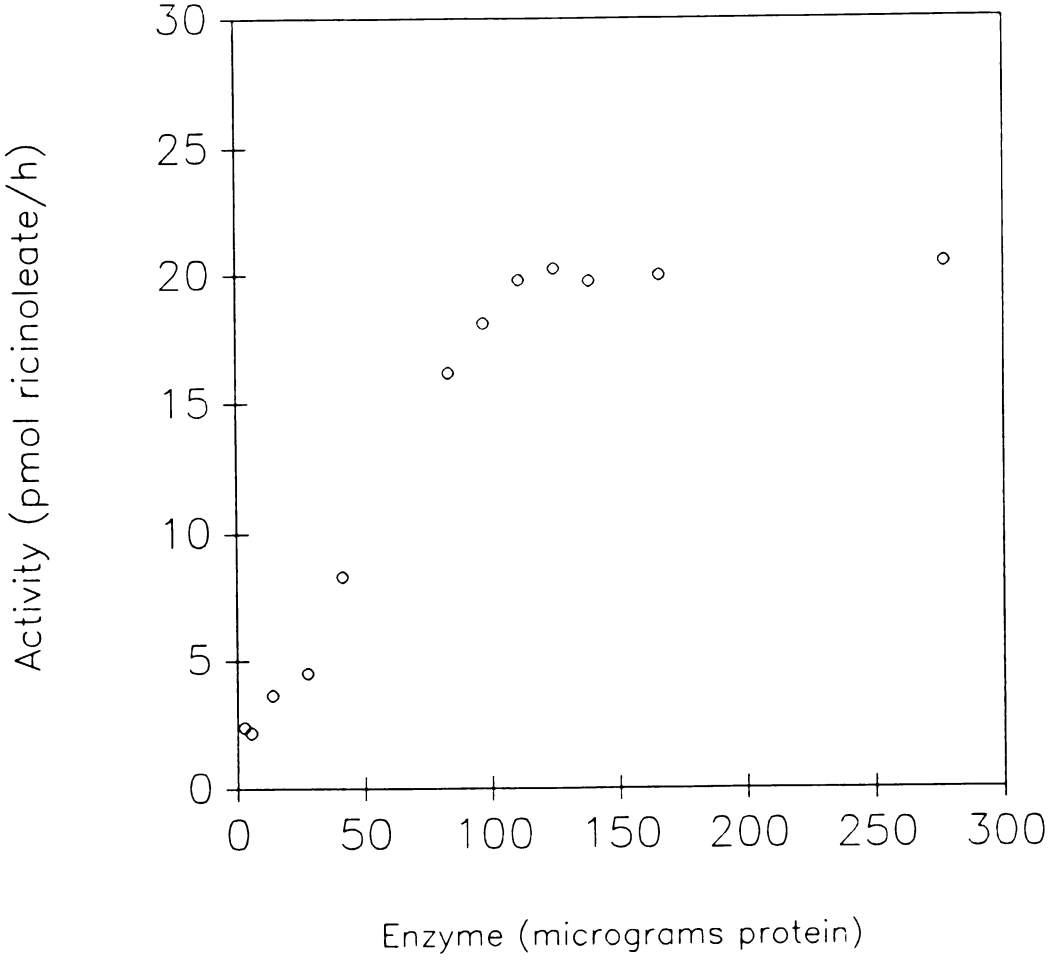


Figure 1. Oleate-12-hydroxylase activity as a function of protein assayed.

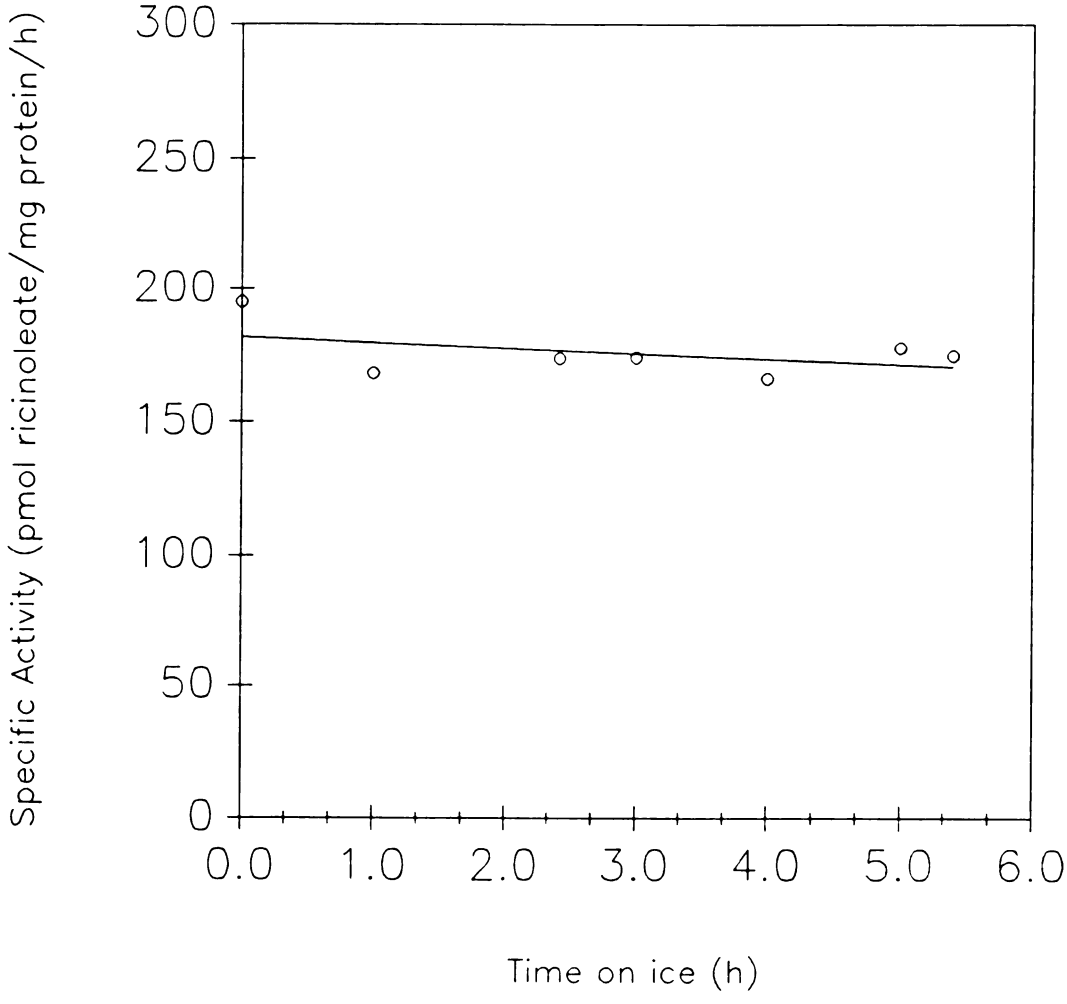


Figure 2. Stability of oleate-12-hydroxylase activity of enzyme incubated on ice.

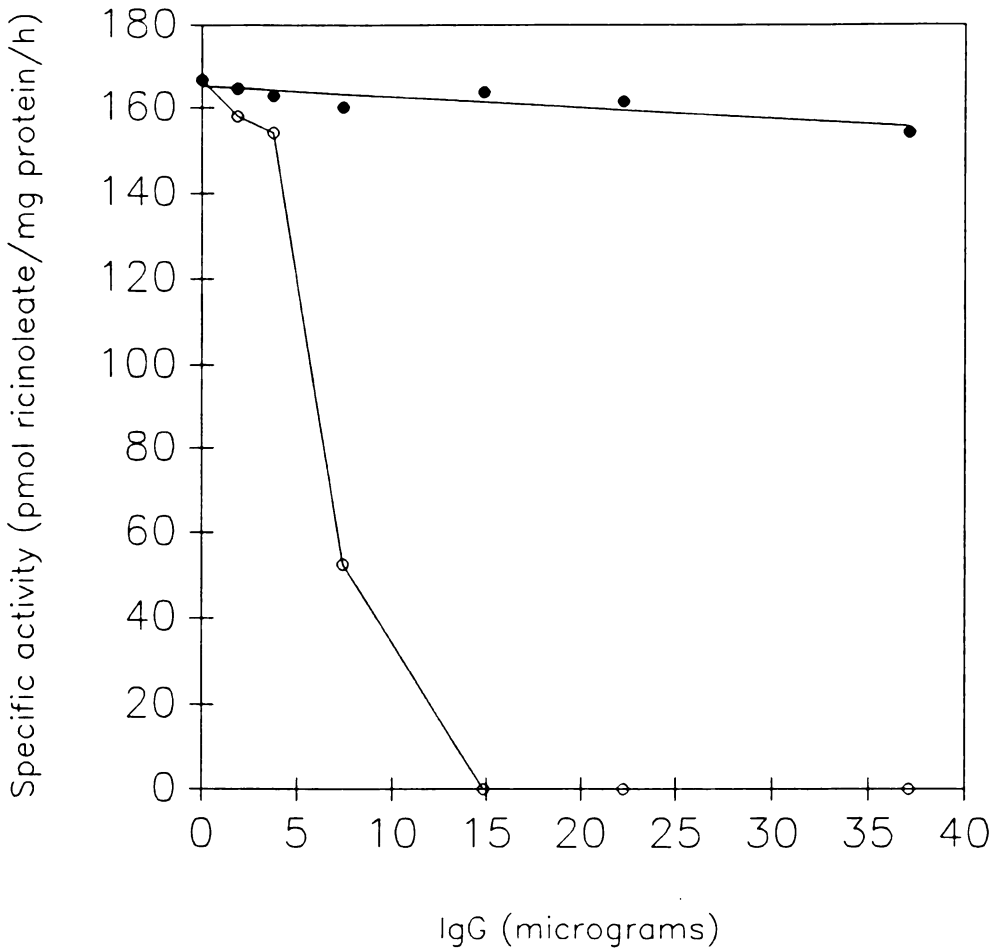


Figure 3. Oleate-12-hydroxylase activity as a function of anti-cytochrome b_5 immunoglobulin (IgG) added. Immunoglobulins were from pre-immune mouse ascites fluid (closed circles) or from a mouse that had been immunized with cauliflower cytochrome b_5 (open circles).

IgG had essentially no effect on oleate-12-hydroxylase activity, anti-cytochrome b_5 IgG strongly inhibited hydroxylation. Complete inhibition was achieved with 15 μg IgG, or an IgG/enzyme protein ratio of 0.18. The same antibody was required at a much higher ratio (4 mg IgG/mg microsomal protein) in order to acheive 93% inhibition of oleate-12-desaturase in microsomal membranes from safflower,[19] although a shorter (2 h) incubation of enzyme and IgG was used in that work.

The assay used here to measure oleate-12-hydroxylase activity employs oleoyl-CoA as substrate, but this may not be the direct substrate of the hydroxylase enzyme. In fact, it has been reported[31] since this work was done that hydroxylation occurs on oleate esterified at *sn*-2 of phosphatidylcholine, as is the case for the microsomal desaturases. Therefore, the assay measures, at least, both acyltransferase activity and hydroxylase activity. It was shown by Kearns *et al.*[19] that the antibodies used in these experiments had no effect on the incorporation of oleic acid into phospholipids, including phosphatidylcholine, in safflower microsomes. Furthermore, the antibodies also block transfer of electrons from cytochrome b_5 to cytochrome *c*, and inhibition of desaturation mediated by the antibodies could be completely quenched by the addition of purified cytochrome b_5.[19] This shows that inhibition of desaturation is due to the involvement of cytochrome b_5 as the electron donor to the safflower microsomal Δ12 desaturase, and by extension, also to the castor oleate-12-hydroxylase. Since the work described here was done, similar results have been reported by Smith *et al.*[32]

PARTIAL PURIFICATION ATTEMPTS

Various strategies for partial purification of oleate-12-hydroxylase activity were investigated in over 30 separate experiments. Since none of these experiments led to any worthwhile purification of the enzyme, they are not presented here in detail. Since it had previously been reported that oleate-12-hydroxylase could be recovered in microsome preparations by differential centrifugation, though often with significant loss of activity,[3,4] attempts were made in these studies to prepare active microsomes.

Preparation of microsomes by standard procedures (removal of cell-debris and fat by low-speed centrifugation followed by centrifugation of the supernatant at 105000 g for 1 h) generally resulted in complete loss of activity, even if the pellet was resuspended in the supernatant. Control enzyme, incubated at the same temperature for the same period of time without ultra-centrifugation, retained full activity. A linear relationship could be established between the amount of activity remaining and the duration (or field strength) of centrifugation. These experiments also revealed an intriguing phenomenon. If an enzyme preparation was partially inactivated by centrifugation, and the pellet resuspended in the supernatant, the quantity of this enzyme sample (μg protein) which saturated the oleate-12-hydroxylase assay was similar to the quantity of the uncentrifuged enzyme (μg protein) required to saturate the oleate-12-hydroxylase assay. For example, if quantities less than 50 μg protein were in the approximately-linear range for the hydroxylase assay, and

quantities greater than 50 μg gave no more activity than did 50 μg (enzyme saturation), then the same was also observed for the centrifuged enzyme, even though the activity of 50 μg was considerably lower.

These experiments suggested that some property of the oleate-12-hydroxylase, such as a requirement for a multi-component complex in a membrane, was destroyed by the force exerted during centrifugation. I am aware of only one other enzyme, an acyl-CoA Δ5-desaturase of meadowfoam (*Limnanthes alba*), for which such a phenomenon has been observed.[33]

Alternatives to differential centrifugation were investigated. Fractionation of membranes in sucrose density-gradients also generally resulted in complete loss of activity, probably due to the extensive centrifugation (83 000 g_{av}, 3 h) required to achieve separation of membrane fractions. In some cases, activity could be recovered from a mixed gradient, but not from any individual fractions of an identical gradient (nor from pair-wise combinations of these fractions). Activity could be retained by briefer centrifugation (such as in Percoll gradients), but at the expense of fractionation, since activity was then recovered throughout the gradient.

Though complete inhibition of the hydroxylase activity by low concentrations of detergents ($<$16 mM CHAPS, $<$1% Triton X-100) suggested that the enzyme was membrane-bound, it could not be assigned to a particular cell fraction on the basis of low-speed centrifugations. After low-speed (e.g. 15 000 g x 15 min) centrifugation, hydroxylase activity was found in both pellet and supernatant fractions.

As an alternative to ultracentrifugation, attempts were made to prepare

microsomes by precipitation with magnesium ions.[34] In this technique, the supernatant from a low-speed centrifugation is adjusted to 50 mM Mg^{2+} and then centrifuged again, microsomes being obtained in the pellet. However, oleate-12-hydroxylase activity pelleted in the second centrifugation independently of the Mg^{2+} concentration added (in the range 0 to 100 mM).

These results generally suggested that the cellular fraction containing oleate-12-hydroxylase activity does not behave independently of other cellular fractions during attempts at fractionation. Possibly, as membrane fraction containing the enzyme adheres non-specifically to various cellular components and therefore gives no simple pattern of fractionation.

Another alternative fractionation method investigated was iso-electric focusing using a Rotofor aparatus (BioRad). No activity could be recovered in any fraction. Precipitations of protein with ammonium sulphate did not give any apparent purification of the activity.

POLYCLONAL ANTIBODIES RAISED AGAINST CRUDE ENZYME EXTRACTS

Because difficulty was experienced in even partial purification of the oleate-12-hydroxylase activity from developing castor seeds, an alternative strategy was investigated for identifying proteins involved in the hydroxylation reaction. It was reasoned that if polyclonal antibodies raised against a complex mixture of proteins,

including the hydroxylase enzyme, could be found to inhibit hydroxylase assays, then a purification of the effective antibody could be attempted. A problem with this approach was the presence in castor seeds of the extremely toxic protein ricin. Two strategies were chosen for raising antibodies in rabbits. In the first strategy, a washed membrane fraction was prepared from seed extracts, and treated with dithiothrietol, to separate the subunits of any remaining ricin, thereby rendering it harmless. Unfortunately, this treatment was not sufficient, due to the extreme lethality of ricin, and this strategy was abandoned and is not described further. The second strategy was to separate SDS-denatured proteins by electrophoresis, then transfer the proteins to a nitrocellulose filter, which was powdered for injection. Such an approach has been used to generate antibodies that inhibit native enzyme activity.[35]

Proteins (3.7 mg) from an enzymically-active oleate-12-hydroxylase preparation were separated on a preparative SDS-polyacrylamide gel, and then transferred to a nitrocellulose membrane. Coomassie-staining of portions of the gel before and after transfer indicated that >50% of the protein was transferred. The membrane was divided into five horizontal strips bearing proteins of the following molecular weight ranges: >95 kD, 69-95 kD, 46-69 kD, 30-46 kD, and 15-30 kD. Protein was fixed to the membrane, then powdered to a form which could be mixed with aqueous buffer to make an injectable suspension. Rabbits, each immunised with protein from one of the molecular weight ranges, provided serum which was used to stain proteins blotted from a gel on which the original enzyme preparation were separated (Figure 4). Antibodies in the sera react principally with proteins of the

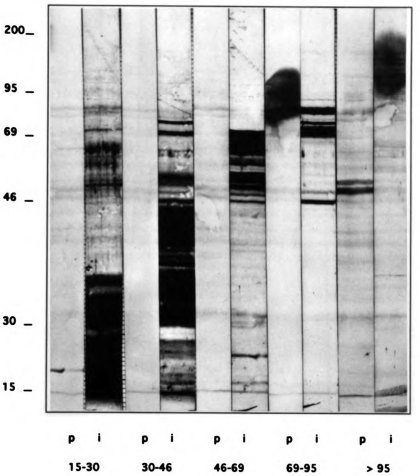


Figure 4. Western blot strips stained with polyclonal antibodies from rabbits immunised with different molecular weight ranges (shown at bottom, kD) of developing castor seed proteins. The same protein preparation was separated on a preparative SDS-PAGE gel, and transferred to give the blot shown. Each strip was developed separately with either preimmune (p) or immune (i) serum. The migration of proteins of given molecular weight (kD) is shown at left.

expected molecular weight range (i.e. that with which the rabbit was immunised), with some cross-reactivity to proteins outside the expected range, possibly due to shared epitopes.

Immunoglobulins from the sera were purified by affinity to protein G, and then incubated (45 min on ice) with castor developing seed extract, before assay of oleate-12-hydroxylase activity. There was no effect of incubation with any of the immunoglobulins on hydroxylase activity (Table 4). In a second experiment, enzyme was incubated with immunoglobulin for a longer period (2 h), and then mixed with a 10% (v/v) slurry of *Stapphylococcus aureus* cells. These bacteria carry a surface protein (protein A) which binds immunoglobulins non-specifically, and can therefore be used to precipitate immunoglobulins and immunoglobulin-protein complexes. The cells were removed after 30 min by centrifugation, and the supernatant was assayed. These incubations also had no effect on hydroxylase activity (Table 4).

Several explanations may account for the lack of effect of the polyclonal antibodies upon enzyme activity. If the hydroxylase is an integral membrane protein, the polyclonal antibodies may have been generated against epitopes buried in the membrane of the non-denatured, membrane-bound enzyme. Such antibodies would not hinder the reaction of the native enzyme. Alternatively, the hydroxylase may be a poor antigen, so that the titre of antibodies specific for the hydroxylase is low. Possibly, other conditions or ratios of antibody to enzyme might be found which would give inhibition of hydroxylase activity. This was not pursued.

Table 4. Oleate-12-hydroxylase activity (pmol ricinoleic acid mg^{-1} protein min^{-1}) following incubation of enzyme (325 μg protein) with immunoglobulins (IgG; 16.6 μg) from rabbits immunised against castor proteins of different molecular weight (MW) ranges. In experiment 1, IgG were incubated with enzyme for 45 min on ice, before assay. In experiment 2, IgG were incubated with enzyme (2 h, on ice) and then precipitated with *Stapphylococcus aureus* (S.A.) cells, before assay of the supernatant.

	Experiment 1		Experiment 2	
	- IgG : 14.6		- IgG, - S.A. : 14.6 - IgG, + S.A. : 15.2	
MW range	**pre-immune**	**immune**	**pre-immune**	**immune**
15-30 kD	14.2	13.8	12.5	11.5
30-46 kD	15.0	14.9	12.2	14.0
46-69 kD	15.3	14.3	12.4	12.5
69-95 kD	13.1	15.0	13.5	12.4
> 95 kD	14.5	15.8	12.2	12.8

DISCUSSION AND CONCLUSIONS

The synthesis of ricinoleic acid involves the stereospecific hydroxylation of oleic acid, found in all plants, at the Δ12 position. Biochemical investigations of this hydroxylation activity have produced very similar results to those of other fatty acid modifying enzymes, namely the microsomal desaturases. All plants have a microsomal oleate desaturase active at the Δ12 position. The substrate of this enzyme[36] and of the hydroxylase[31] appears to be oleate esterified to the *sn-2* position of phosphatidylcholine. The modification occurs at the same position (Δ12) in the carbon chain, and requires the same cofactors, namely electrons from NADH via cytochrome b_5 (Figure 3, references 19,32) and molecular oxygen (Appendix). Neither enzyme is inhibited by carbon monoxide,[3,4] the characteristic inhibitor of cytochrome P450 enzymes.

This last observation is in contrast to another set of plant fatty acid modifying enzymes studied, namely the hydroxylases and epoxidases involved in cutin biosynthesis. These enzymes act at the terminal positions (ω, ω-1) of the carbon chain and are characteristically P450 enzymes. The possible evolutionary origin of the castor oleate-12-hydroxylase is discussed further in chapter 4. It is merely observed at this point that the greater similarities to the microsomal destaturases make these the more likely progenitors of the castor seed hydroxylase than the hydroxylases of cutin biosynthesis. The opposite relationship may apply for another fatty acid oxidase, responsible for linoleate 12,13-epoxidation in *Euphorbia lagascae*, which has

the characteristics of a cytochrome P450.[37]

The studies reported in this chapter provided new information relating to the developmental pattern of oleate-12-hydroxylase activity in developing seeds, and demonstrate the involvement of cytochrome b_5 and molecular oxygen in hydroxylation. However, little direct progress was made toward the stated goal of cloning a gene(s) encoding the hydroxylase. Attempts at partial purification of the enzyme indicate that this approach would be very difficult. I am not aware of the purification of any membrane protein from an oilseed. Difficulty in purification has also been encountered for microsomal desaturases, leading to attempts to clone genes by genetic approaches, such as chromosome walking[38]. Therefore, the next step toward the goal was also to take a genetic approach (chapter 3).

REFERENCES

1. James, A.T. Hadaway, H.C., and Webb, J.P.W., The biosynthesis of ricinoleic acid, *Biochem. J.*, 95, 448-452, 1965.

2. Morris, L.J., The mechanism of ricinoleic acid biosynthesis in Ricinus communis seeds, *Biochem. Biophys. Res. Commun.*, 29, 311, 1967.

3. Galliard, T., and Stumpf, P.K., Fat metabolism in higher plants XXX. Enzymatic synthesis of ricinoleic acid by a microsomal preparation from developing Ricinus communis seeds, *J. Biol. Chem.*, 241, 5806, 1966.

4. Moreau, R.A., and Stumpf, P.K., Recent studies of the enzymic synthesis of ricinoleic acid by developing castor beans, *Plant Physiol.*, 67, 672, 1981.

5. Canvin, D.T., The biosynthesis of long-chain fatty acids in the developing castor bean, *Can. J. Bot.*, 43, 49-62, 1965.

6. Howling, D., Morris, L.J., Gurr, M.I., and James, A.T., The specificity of fatty acid desaturases and hydroxylases. The dehydrogenation and hydroxylation of monoenoic acids, *Biochim. Biophys. Acta*, 260, 10, 1972.

7. Křen, V., Řezanka, T., and Řeháček, Z., Occurrence of ricinoleic acid in submerged cultures of various *Claviceps* sp., *Experentia*, 41, 1476-1477, 1985.

8. Morris, L.J., Hall, S.W., and James, A.T., The biosynthesis of ricinoleic acid by *Claviceps purpurea*, *Biochem. J.*, 100, 29c-30c, 1966.

9. Morris, L.J., Mechanisms and sterochemistry in fatty acid metabolism, *Biochem. J.*, 118, 681-693, 1970.

10. Ahmad, F., Schiller, H., and Mukherjee, K.D., Lipids containing isoricinoleoyl (9-hydroxy-cis-12-octadecenoyl) moieties in seeds of Wrightia species, *Lipids*, 21, 486, 1986.

11. Ahmad, F., and Mukherjee, K.D., Biosynthesis of lipids containing isoricinoleic (9-hydroxy-cis-12-octadecenoic) acid in seeds of Wrightia species, *Z. Naturforsch.*, 43c, 505, 1988.

12. Greenwood, J.S., and Bewley, J.D., Seed development in *Ricinus communis* (castor bean). I. Descriptive morphology, *Can. J. Bot.*, 60, 1751-1760, 1982.

13. Bradford, M.M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of dye-binding, *Anal. Biochem.*, 72, 248-254, 1976.

14. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*, 227, 680-685, 1970.

15. Towbin, H., Staehelin, T., and Gordon, J., Electrophoretic transfer of proteins from acrylamide gels to nitrocellulose sheets: procedure and some applications, *Proc. Natl. Acad. Sci. USA*, 76, 4350-4354, 1979.

16. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

17. Diano, M., Bivic, A. le, and Hirn, M., A method for the production of highly specific polyclonal antibodies, *Anal. Biochem.*, 166, 224-229, 1987.

18. Hurn, B.A.L., and Chantler, S.M., Production of reagent antibodies, *Meth.*

Enzymol., 70, 104-142, 1980.

19. Kearns, E.V., Hugly, S., and Somerville, C.R., The role of cytochrome b_5 in Δ12 desaturation of oleic acid by microsomes of safflower (*Carthamus tinctorius* L.), *Arch. Biochem. Biophys.*, 284, 431-436, 1991.

20. Kessler, S.W., Cell membrane antigen isolation with the staphylococcal protein A-antibody adsorbent, *J. Immunol.*, 117, 1482-1490, 1976.

21. Greenwood, J.S., and Bewley, J.D., Seed development in *Ricinus communis* cv. Hale (castor bean). II. Accumulation of phytic acid in the developing endosperm and embryo in relation to the deposition of lipid, protein, and phosphorus, *Can. J. Bot.*, 62, 255-261, 1984.

22. Oshino, N., Imai, Y., and Sato, R., A function of cytochrome b_5 in fatty acid desaturation by rat liver microsomes, *J. Biochem. (Tokyo)*, 69, 155-167, 1971.

23. Strittmatter, P., Spatz, L., Corcoran, D., Rogers, M.J., Setlow, B., and Redline, R., Purification and properties of rat liver microsomal stearoyl coenzyme A desaturase, *Proc. Natl. Acad. Sci. USA*, 71, 4565-4569, 1974.

24. Ferrante, G., and Kates, M., Characteristics of the oleoyl- and linoleoyl-CoA desaturase and hydroxylase systems in cell fractions from soybean cell suspension cultures, *Biochim. Biophys. Acta*, 876, 429, 1986.

25. Salaun, J.-P., Weissbart, D., Durst, F., Pflieger, P., and Mioskowski, C., Epoxidation of cis aand trans Δ^9-unsaturated lauric acids by a cytochrome P-450-dependent system from higher plant microsomes, *FEBS Lett.*, 246, 120, 1989.

26. Blee, E., Schuber, F., Stereochemistry of the epoxidation of fatty acids catalyzed by soybean peroxygenase, *Biochem. Biophys. Res. Commun.*, 173, 1354, 1990.

27. Grechkin, A.N., Kukhtina, N.V., Gafarova, T.E., and Kuramshin, R.A., Oxidation of [1-^{14}C]linoleic acid in isolated microsomes from pea leaves, *Plant Sci.*, 70, 175, 1990.

28. Janistyn, B., Evidence for conversion of arachidonic acid to hydroxyeicosatetraenoic acids by a cell-free homogenate of maize seedlings, *Phytochemistry*, 29, 2453, 1990.

29. Fahlstadius, P., Absolute configurations of 9,10-epoxydodecanoic acids

biosynthesized by microsomes from jerusalem artichoke tubers, *Phytochemistry*, 30, 1905, 1991.

30. von Wettstein-Knowles, P.M., Waxes, cutin, and suberin, in *Lipid Metabolism in Plants*, T.S. Moore Jr., Ed., CRC Press, Boca Raton, 1993, 127-166.

31. Bafor, M., Smith, M.A., Jonsson, L., Stobart, K., and Stymne, S., Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (Ricinus communis) endosperm, *Biochem. J.*, 280, 507, 1991.

32. Smith, M.A., Jonsson, L., Stymne, S., and Stobart, K., Evidence for cytochrome b_5 as an electron donor in ricinoleic acid biosynthesis in microsomal preparations from developing castor bean (*Ricinus communis* L.), *Biochem. J.*, 287, 141-144, 1992).

33. Moreau, R.A., Pollard, M.R., and Stumpf, P.K., Properties of a Δ5-fatty acyl-CoA desaturase in the cotyledons of developing *Limnanthes alba*, *Arch. Biochem. Biophys.*, 209, 376-384, 1981.

34. Diesberger, H., Muller, C.R., and Sanderman, H., Jr., Rapid isolation of a plant microsomal fraction by Mg^{2+}-precipitation, *FEBS Lett.*, 43, 155-158, 1974.

35. Christie, P.J., Ward, J.E., Gordon, M.P., and Nester, E.W., A gene required for transfer of T-DNA to plants encodes an ATPase with autophosphorylating activity, *Proc. Natl. Acad. Sci. USA*, 86, 9677-9681, 1989.

36. Heinz, E., Biosynthesis of polyunsaturated fatty acids, in *Lipid Metabolism in Plants*, T.S. Moore Jr., Ed., CRC Press, Boca Raton, 1993, 33-89.

37. Bafor, M., Smith, M.A., Jonsson, L., Stobart, K., and Stymne, S., Biosynthesis of vernoleate (cis-12-epoxyoctadec-cis-9-enoate) in microsomal preparations from developing endosperm of Euphorbia lagascae, *Arch. Biochem. Biophys.*, 303, 145-151, 1993.

38. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M., and Somerville, C.R., Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*, *Science*, 258, 1353-1355, 1992.

CHAPTER 3

GENETIC EXPERIMENTS

ABSTRACT

A genetic approach to isolating a gene encoding oleate-12-hydroxylase is described. A cDNA library was prepared from developing castor endosperm, in a yeast expression vector. Clones of the library were expressed in yeast cultures, which were analysed by gas chromatography for fatty acid composition. Two methods were used to pool and analyse cultures for the appearance of ricinoleic acid or other changes in fatty acid composition. In experiment 1, five cultures were pooled per analysis, and of 2300 clones screened, none showed differences in fatty acid composition. In experiment 2, 2500-3000 clones were analysed as pools of 500, with enrichment of polar fatty acids prior to gas chromatography. No ricinoleic acid was detected.

It is suggested that some problem related to heterologous expression or the nature of the oleate-12-hydroxylase precluded detection of its expression in this experiment.

INTRODUCTION

Since the biochemical experiments described in chapter 2 were not productive towards my overall goal of gaining a better molecular understanding of the oleate-12-hydroxylase, I decided to employ a different strategy. One genetic approach to cloning a gene for oleate-12-hydroxylase is the subject of this chapter. I sought to identify a *Ricinus communis* cDNA which when expressed in growing yeast cells would be actively transcribed and translated, leading to enzyme activity which could be detected by the accummulation of ricinoleic acid. The rationale for this experiment is described below.

Expression of Heterologous Enzymes in Yeast

Expression of active cytosolic enzymes from eukaryotes has been possible in *E. coli*, but expression, targetting and processing of eukaryotic proteins normally found in other compartments of the cell requires expression in another eukaryote, such as yeast. In many cases, proteins are targetted in the yeast cell in a very similar manner to that in the native cell. Mammalian proteins have been functionally expressed in yeast cells. For example, cytochrome P450 laurate ω-1 hydroxylase from rabbit accumulates in yeast microsomes in an active form when the rabbit cDNA is expressed in yeast cells.[1-4] Similar results have also been obtained with cytochrome P450MC from rat liver.[5] Likewise, a mouse liver cytochrome P450 cDNA expressed in yeast was integrated into the microsomal membrane in a fully functional form,

interacting with the endogenous cytochrome P450 reductase.[6] Translation and proper targetting of heterologous proteins is, however, not uniformly the case. When cDNAs for the four subunits of the nicotinic acetylcholine receptor from electric ray were transformed into yeast, all were transcribed, but only two of the proteins accumulated.[7] For the other two subunits it was necessary to replace the original signal sequences with that of a yeast gene.

Plant cDNAs have also been functionally expressed and targetted in yeast. Barley α-amylases 1 and 2 were efficiently secreted into the culture medium under the direction of their own signal peptides.[8] Similarly, when the bean vacuolar protein phytohemagglutinin was expressed in yeast, it was efficiently targetted to the yeast vacuole.[9]

In an example particularly relevant to this work, single copy expression of the rat stearoyl-CoA desaturase gene in a yeast mutant (*ole1*) deficient in the homologous enzyme, complements the mutant.[10] These rat and yeast stearoyl-CoA desaturases are microsomal proteins which form part of a three-component system involving NADH-cytochrome b_5 reductase, cytochrome b_5, and the desaturase. The functional replacement of the yeast gene by the transformed rat gene therefore demonstrates that the rat gene product is correctly targetted to the microsomes and can functionally interact with yeast cytochrome b_5. This is in spite of the fact that the rat and yeast desaturase proteins share only 36% identity, suggesting that aspects of these proteins involved in cytochrome b_5 interaction and cytochrome b_5 itself are particularly conserved. Alternatively, no specific sequence is required for the interaction. Since

the experiments of this chapter were done, expression in yeast of various plant proteins has been reported, including functional expression of a plant cytochrome P450 (see Discussion and Conclusions).

Availability of Substrates and Cofactors

For functional expression, the plant enzyme must have access to all necessary substrates and cofactors. It is clear that NADH and oxygen are present in yeast, and it seems at least quite possible that the yeast NADH-cytochrome b_5/cytochrome b_5 system will be able to provide electrons to castor oleate-12-hydroxylase as they do to the rat stearoyl-CoA desaturase. The true substrate of the oleate-12-hydroxylase was not known at the time of this experiment, but was thought to be either oleoyl-CoA (which is the substrate supplied to assays of the enzyme in castor endosperm extracts), or oleate esterified to phoshatidylcholine. Oleate is a major component of yeast membranes, and phosphatidylcholine is the major phospholipid of yeast microsomes,[11] so it appears likely that both oleoyl-CoA and phosphatidylcholine-esterified oleate should be available at the yeast microsome. It was considered that pH, other ionic and solute differences between castor and yeast microsomal environments should be relatively subtle and not sufficient to abolish activity of the plant enzyme in yeast.

Single-Gene Assumption

One fundamental assumption of this experiment was that a single castor cDNA

is sufficient to give oleate-12-hydroxylase activity. There is no evidence available for or against this assumption, so that for this reason in particular, this experiment was viewed as risky.

Level of Accumulation of Ricinoleate

Ricinoleate produced by the enzyme expressed in yeast must accumulate to detectable amounts. Since ricinoleate is an unusually polar acyl group, and is found only in the triglyceride fraction of castor lipids,[12,13] it was thought that high levels of accumulation in yeast might be disruptive to membrane function and were, therefore, not to be expected. In experiment 1, total fatty acids of the yeast cultures were analysed for the presence of ricinoleic acid. It was estimated that ricinoleic acid could be detected if it constituted approximately 0.1-0.5% of the fatty acids of a particular culture.

Possibility of Further Metabolism of Ricinoleate

For ricinoleate to accumulate to detectable levels, it must also not be metabolised further, or the metabolite should also be detectable. There is a report[14] that hydroxystearic acids support the growth of yeast in anaerobic condtions when they are not otherwise able to grow, due to their inability to make unsaturated fatty acids (similar to experiments reviewed in chapter 1). This appeared to be due to acetylation of the hydroxyl group and use of the acetoxy acids as substitutes for oleic acid in growth. The acetylase activity was also active towards ricinoleic acid. Thus

synthesis of ricinoleic acid in yeast may be followed by its acetylation or other modification. Chromatograms were therefore analysed not only for a peak corresponding to ricinoleate, but also for any other novel peaks.

Number of Clones to be Screened

All the above conditions being met, the success of the experiment would depend upon being able to screen a sufficient number of clones, based upon the expected frequency of the oleate-12-hydroxylase cDNA in the library used. The library was analysed for the freqency of stearoyl-ACP desaturase[15] clones, to use as an approximation for the frequency of clones for oleate-12-hydroxylase. This assumption of similar frequencies is made on the basis that both are involved in the same pathway (i.e. producing ricinoleate for triglyceride synthesis) and that both enzymes accomplish fatty acid oxidations, therefore being required to overcome the same high energy of carbon-hydrogen bond cleavage, and so might have similar turnover numbers (discussed further in chapter 4). Since the library was constructed by a non-directional approach, on average only half the clones encoding the pututative hydroxylase will be in the correct orientation for expression. Furthermore, a percentage of clones will not contain a full-length copy of the coding sequence, causing some further increase in the number which need be analysed.

MATERIALS AND METHODS

Construction of cDNA Library

Total RNA was purified from developing stage III to stage V[16] cellular endosperm plus embryo by the technique of Puissant & Houdebine,[17] with the following modifications: tissue was ground under liquid nitrogen instead of at 4°C, and both the LiCl and $CHCl_3$ extractions were repeated once. Poly(A)$^+$ RNA was enriched by two rounds of chromatography on oligo dT cellulose[18] and analysed by electrophoresis through formaldehyde-containing agarose gels.[18] The final poly(A)$^+$ RNA still contained obvious ribosomal RNA bands but these were less prominent compared to the background smear than in the total RNA. Complementary DNA was prepared using a kit ("Librarian IV", Invitrogen) according to the instructions of the manufacturer. First strand cDNA (1.65 μg) was synthesised from poly(A)$^+$ RNA (5 μg) by priming with oligo dT and extension by avian myeloma virus reverse transcriptase. The RNA was nicked by *E. coli* RNaseH, forming primers for second-strand cDNA synthesis by *E. coli* DNA polymerase I. Any nicks in the dsDNA were repaired with *E. coli* DNA ligase. Ends of the dsDNA were made blunt with T4 DNA polymerase for ligation of *Bst*X1 non-palindromic linkers. The cDNA was size-selected by agarose-gel electrophoresis, and molecules larger than ~750 bp were ligated into the *Bst*X1-digested pYES2.0 vector (Figure 5) and transformed into *E. coli* strain INV1αF', yielding four pools containing a total of 1.42 X 10^6 transformants.

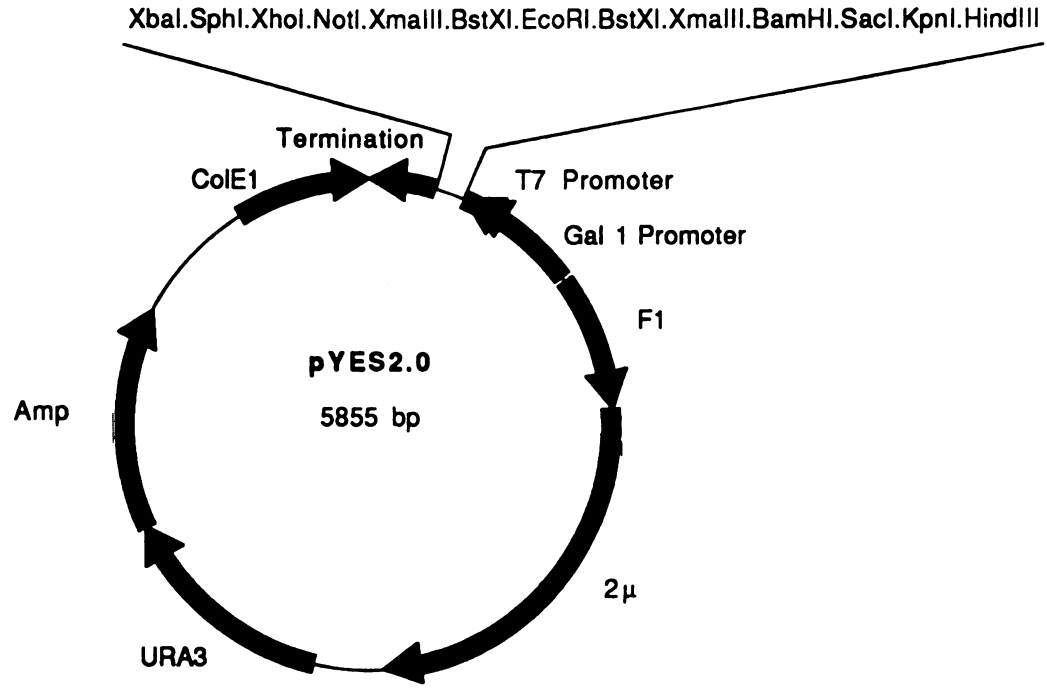


Figure 5. The vector pYES2.0 (Invitrogen) used for expression of cDNA in yeast. Castor seed cDNA was ligated at the *Bst*XI sites.

Screening of Library for Stearoyl-ACP Desaturase

Three plates of the pYES2.0 library were grown until small colonies (53 000 total) were visible. A nitrocellulose filter (Schleicher & Schüll BA85) was laid on each plate, its position marked, and lifted off to a fresh plate, the adhering colonies now facing upwards. Care was necessary that both plate and filter were not too moist, to avoid smearing of the colonies. The original plate was incubated for a few h at 37°C to recover colonies, while the filters were processed as folllows. Each filter was sequentially placed, colony side up, on Whatman 3MM paper moist with 10% SDS (3 min), denaturing solution (0.5 M NaOH, 1.5 M NaCl; 5 min), neutralising solution (0.5 M Tris-Cl pH 7.4, 1.5 M NaCl; 5 min), and 2 X SSC (0.3 M NaCl, 0.03 M Na-Citrate, pH 7.0). The filters were then air-dried for ca. 1 min before pressing twice between sheets of filter paper to remove cell debris. After air-drying a further 30 min, DNA was fixed to the filters by baking *in vacuo* at 80°C for 1-2 h.

The filters were prehybridised in a minimal volume of 0.25% nonfat dry milk, 6 X SSC, 10% dextran sulphate, at 68°C, before addition of the probe and hybridisation overnight. Probes were labelled by random priming[18] and purified of unincorporated nucleotides by ethanol precipitation in the presence of ammonium acetate.[18] The filters were washed three times in 2 X SSC, 0.1% SDS at 68°C, then exposed to X-ray film. Hybridizing colonies were picked and replated for a second round of hybridization yielding positive clones.

Transformation of Yeast

DNA of the four pools of the library was prepared from 300 ml *E. coli* cultures by an alkaline lysis procedure[18] and purified by banding in CsCl gradients. This DNA was transformed into yeast by electroporation according to Becker & Guarente.[19]

Selective Yeast Growth Media

Media for selective growth of yeast contained (per litre):

6.7 g yeast nitrogen base without amino acids (Difco)

10 ml amino acid stock solution

amino acid stock solution (per 200 ml):

8 g casamino acids (Difco)

0.4 g adenine sulphate

0.6 g L-leucine

0.4 g L-tryptophan

0.6 g L-lysine-HCl

0.6 g L-histidine-HCl

Media contained 2% carbon source (glucose, galactose, or sodium lactate) added after autoclaving. Media for plates were solidified with 2% Dico Bacto Agar.

β-Galactosidase Assays

β-galactosidase activity was assayed by measuring *o*-nitrophenol hydrolysed from ONPG (*o*-nitrophenyl-β(D)-galctopyranoside). Absorbance (λ=600 nm) of the yeast culture was measured and duplicate samples (1 ml) were pelleted in a microfuge (5 min). Pellets were washed with 1 ml Z-buffer (0.06 M Na_2HPO_4, 0.04 M NaH_2PO_4, 0.01 M KCl, 0.001 M $MgSO_4$, adjusted to pH 7.0 with NaOH, and 0.03 M β-mercaptoethanol added just prior to use). Z-buffer (950 μl) was added to the pellet, followed by 2 drops chloroform (from a 9-inch pasteur pipette) and 2 drops 0.1% SDS. The cells were vortexed (10 s) at maximum speed. 200 μl ONPG (4 mg ml^{-1} aqueous solution) at 30°C was added at a recorded time, vortexed, and incubated at 30°C. The time was recorded when the reaction turned yellow, and 0.5 ml 1 M Na_2CO_3 added and vortexed to stop the reaction. The tubes were microcentrifuged 2 min, and absorbance (λ=420 nm) of the supernatant measured spectrophotometrically. Results (nmol *o*-nitrophenol/A_{600}/min = A_{420} x 79.812/(A_{600} x time)) were calculated using an extinction coefficient for *o*-nitrophenol of 21300 $M^{-1}cm^{-1}$.

Gas Chromatography of Yeast Fatty Acid Methyl Esters (FAMEs)

For experiment 1, yeast cells from five 5 ml cultures were pooled, pelleted (SS-34 rotor, 10 000 rpm, 5 min), resuspended in methanol (2 x 0.5 ml) and

transferred to a glass scew-cap tube. 1.5 M methanolic HCl (2 ml) was added, the tube capped with a teflon-lined cap, and heated to 80°C for 1 h. Upon cooling, 2 ml hexane:isopropanol (3:2) and 1 ml 0.2 M Na_2SO_4 were added and the FAMEs removed in the hexane phase, which could be stored at -20°C. The samples were dried under a stream of nitrogen, redissolved in hexane (100 μl) and transferred to a gas chromatograph (GC) vial. Any hydroxyl groups were derivatized at room temperature by addition to the vial of 1 μl trimethylsilylimidazole reagent (TriSil Z, Pierce). Some samples were halved and one aliquot spiked with authentic methyl-ricinoleate (4 μg, Sigma).

For experiment 2, the cell pellet was thawed in 5 ml 15% methanolic KOH and saponified at 80°C for 1 h. The fatty acids were neutralized with 2.5 ml 6 M HCl, then extracted by addition of 1 ml 1 M NaCl and 3 ml hexane. The hexane phase was removed and the aqueous phase further extracted with hexane (2 x 1 ml). The pooled hexane phases were dried under a stream of nitrogen and redissolved in chloroform. A slurry (1:1) of silicic acid (BioSilA) in chloroform was added (1 ml), then the solvent removed under nitrogen. The fatty-acid containing silicic acid powder was transferred to a small column and eluted with 4 ml chloroform (neutral fraction) followed by 2 ml methanol (polar fraction). These fractions were dried under nitrogen and redissolved in methanolic HCl (1 M) and heated at 80°C for 1 h. The methyl esters were extracted into hexane, dried down, redissolved in a small volume (100 μl) of anhydrous hexane, and transferred to a GC vial. After GC analysis of these FAMEs, any hydroxyl groups were derivatized by the addition of

trimethylsilylimidazole and analysed by GC again, comparing the chromatograms of underivatized and derivatized samples.

Gas chromatography was done on a Hewlet-Packard 5890 series II instrument using SP-2330 glass capillary columns (30 m long, 0.75 mm internal diameter, 0.20 μm film thickness) with helium as carrier gas.

EXPERIMENTAL DESIGN

Complementary DNAs expressed in developing castor endosperm during active oleate hydroxylation were cloned into a plasmid vector suitable for expression of the cDNAs in yeast (*Saccharomyces cerevisiae*). The plasmid chosen (pYES2.0) had several important features (Figure 5). Firstly, it has a ColE1 origin of replication and a β-lactamase gene allowing maintenance and selection in *E. coli*, respectively. Secondly, it contains the 2μ origin of yeast plasmids which confers high copy levels in yeast cells and thus favours high levels of expression of the insert cDNA. Selection in yeast depends upon complementation of uracil auxotrophy in the host strain by the plasmid *URA3* gene. Transcription of the insert cDNA is under the control of the GAL1 promotor in GAL^+ host strains, so that transcription can be repressed by growth on glucose, or activated by growth on galactose in the absence of glucose. This particular feature was considered desirable in the possibility that the inserted cDNA would be toxic to yeast cells when expressed.

The yeast host strain CGY2557 (derivative of AB1380 x X15-3A from

Douglas Smith, Collaborative Research, Inc., Bedford, MA. *MATα*, GAL$^+$, *ura3-52, leu2-3, trp1, ade2-1, lys2-1, his5, can1-100*) was chosen for the experiment on the basis of its GAL$^+$ phenotype and uracil auxotrophy.

Since a novel peak was sought in gas chromatograms, it was possible to pool a number of yeast cultures for each analysis, thus increasing the overall number of clones that could be analysed. In experiment 1, five cultures were pooled for each analysis. The number of cultures pooled was greatly increased in experiment 2, based on the reconstruction experiment described below.

PRELIMINARY EXPERIMENTS

Testing GAL Phenotype of Yeast Host Strain

A preliminary control experiment was done to test the GAL$^+$ phenotype of the the yeast strain CGY2557. This is essentially a test of its ability to grow on galactose and induce the GAL1 promoter. The plasmid pCGS286 was obtained from Dr Susan Gibson and has a GAL1/lacZ fusion under the control of the GAL1 promoter. A random clone from the library constructed in pYES2.0 and pCGS286 were used to transform CGY2557 to uracil prototrophy. Transformants were inoculated into selective medium with 2% lactate as sole (GAL-promoter neutral) carbon source. These cultures were grown for 3 days at 30°C until visibly turbid (lactate supports low growth rates), and then supplemented to either 0.2% glucose or 0.2% galactose.

Cultures were harvested and assayed for β-galactosidase activity at the time of sugar addition and on subsequent days. Results (Table 5) demonstrate that β-galactosidase activity was detectable only in pCGS286-bearing cells, and only when these were supplemented with galactose. The highest β-galactosidase activities were attained within one day, declining in subsequent days. Thus induction of the GAL1 promoter occurs under the conditions used in these experiments.

Reconstruction Study for Experiment 2

In experiment 2, larger pools of cultures were analysed simultaneously to facilitate the analysis of a greater number of clones. To achieve this, the samples were first enriched for polar fatty acids before analysis. A preliminary reconstruction experiment was therefore done to verify the sensitivity of the techniques used, and to give an estimate of the content of ricinoleic acid which might be detected. Soybean phoshatidylcholine (50 mg) was dissolved in a minimal volume of chloroform, then saponified with 2 ml 15% methanolic KOH at 80°C for 30 min. The fatty acids were extracted into hexane, and different samples were spiked with 0-100 μg methyl-ricinoleate, then dried under a stream of nitrogen and redissolved in chloroform. A slurry (1:1) of silicic acid (BioSilA) in chloroform was added (1 ml), then the solvent removed under nitrogen. The fatty-acid containing silicic acid powder was transferred to a small column and eluted with 3 ml chloroform (neutral fraction) followed by 2 ml methanol (polar fraction). These fractions were dried under

Table 5. β-galactosidase activities of cultures of CGY2557 harbouring a random clone (pYESr) or pCGS286, supplemented with either glucose or galactose, at the time of sugar addition (a) and on subsequent days. All values are means of duplicate assays of 5 cultures (n=10), except where indicated (*n=8).

	pYESr		pCGS286	
Time	**Glucose**	**Galactose**	**Glucose**	**Galactose**
		units of activity		
a	0.00983		0.0781	
a+1d	0.0193	0.0205	0.0229	10.35
a+2d	0.0134	0.0124	0.0364	5.97
a+3d	0.0169	0.0167	0.0828	3.95*
a+4d	0.0256	0.0241	0.138	3.62

nitrogen and redissolved in methanolic HCl (1 M) and heated at 80°C for 1h. The methyl esters were extracted into hexane, dried down, redissolved in a small volume (100 μl) of anhydrous hexane, and hydroxyl groups were derivatized by the addition of trimethylsilylimidazole. These samples were analysed by gas chromatography for TMS-methyl-ricinoleate. TMS-methyl-ricinoleate could only be detected in the polar fraction from samples originally spiked with 10 μg or 100 μg methyl ricinoleate. This indicated that if the original sample (36 mg fatty acids) represented the fatty acids extracted from 500 yeast cultures, ricinoleate could be detected if it comprised approximately 10% of the total fatty acids of an individual culture.

RESULTS

Experiment 1

Yeast transformants from the cDNA library constructed from developing castor endosperm RNA were ordered into 96-well plates. Each clone could then be used to inoculate a 5 ml culture, which was pooled with four other cultures and analysed by gas chromatography for any changes in the fatty acid methyl ester (FAME) profile. If a gas chromatogram appeared different from normal, fresh cultures were grown of the five clones comprising that pool, and each culture analysed individually. If there was a true difference caused by the cDNA expressed in one clone, then the change in the FAME profile was expected to have a five-fold greater magnitude than that observed in the analysis of the pool. This could be confirmed by comparing growth in glucose-

and galactose-supplemented medium; expression of the cDNA is repressed by glucose, hence a change in the FAME profile caused by expression of the cDNA should only be observed when grown with galactose in the absence of glucose.

An example of a normal gas chromatogram from a primary screen is shown in Figure 6. Periodically, samples were halved, and one aliquot spiked with methyl ricinoleate to verify that the derivatisation reaction was working satisfactorily, and to recalibrate the elution time of TMS-methyl ricinoleate. The sample shown in Figure 6 was such a sample; the spiked chromatogram is shown in Figure 7. An example of a chromatogram in which there is a putative difference in the FAME profile and which was selected for rescreening is shown in Figure 8.

A total of 2300 clones (460 pools) were analysed by this method. No repeatable differences in the FAME profile were found. Thus, of 2300 cDNA clones, none had a detectable effect on fatty acid metabolism of the yeast cell. This may be due (for various reasons discussed in Introduction) to problems associated with heterologous expression of randomly chosen cDNA clones, or due to the absence from this sample of 2300 clones of any which might be involved in fatty acid metabolism (e.g. a fatty acid desaturase or hydroxylase). It was desirable to screen a larger number of clones, and so a procedure was devised which allowed larger pools to be analysed, in experiment 2.

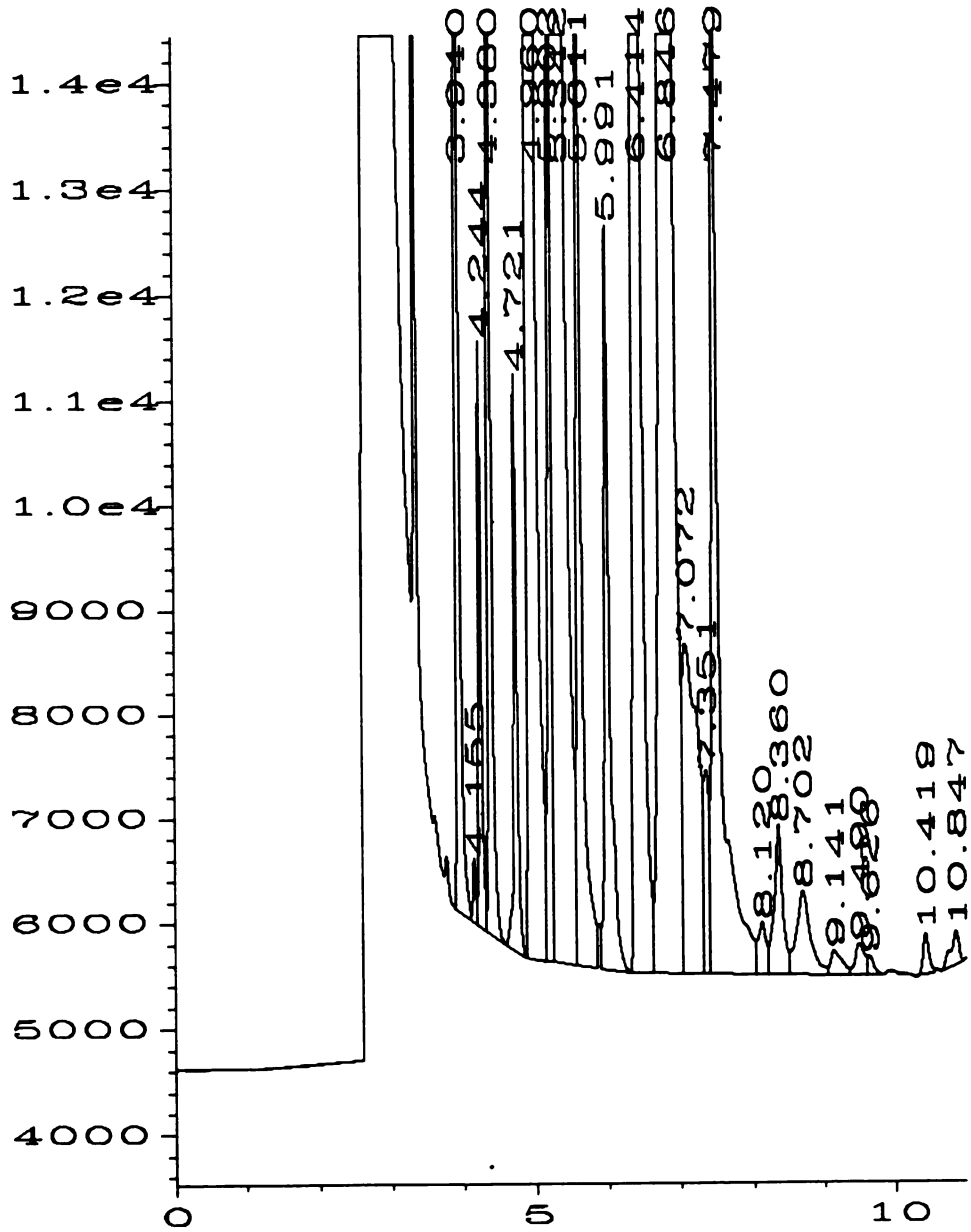


Figure 6. Gas chromatogram of fatty acid methyl esters from sample 145 (cultures 8H5 through 8H9). An attenuation was used which favoured visual inspection for small peaks. Vertical axis: units of signal strength (flame ionization detector); horizontal axis: retention time (min after injection). Elution times of abundant yeast FAMEs in this chromatogram were: 16:0 4.960 min, 16:1 5.342 min, 18:0 6.414 min, 18:1 6.846 min.

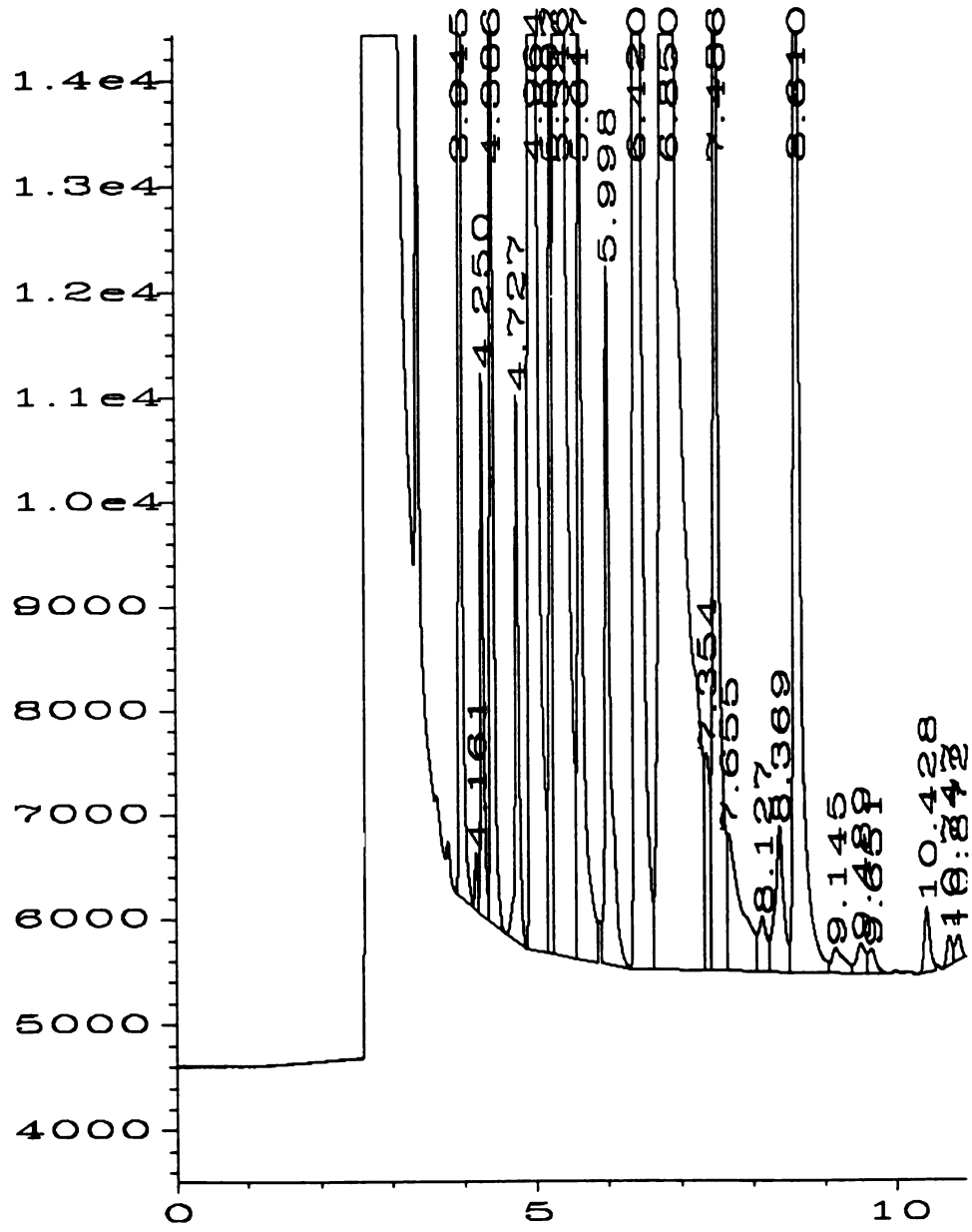


Figure 7. Gas chromatogram of fatty acid methyl esters from sample 145 (cultures 8H5 through 8H9), spiked with authentic methyl-ricinoleate. Note the elution of TMS-methyl ricinoleate at 8.610 min. (Attenuation and axes as for Figure 6).

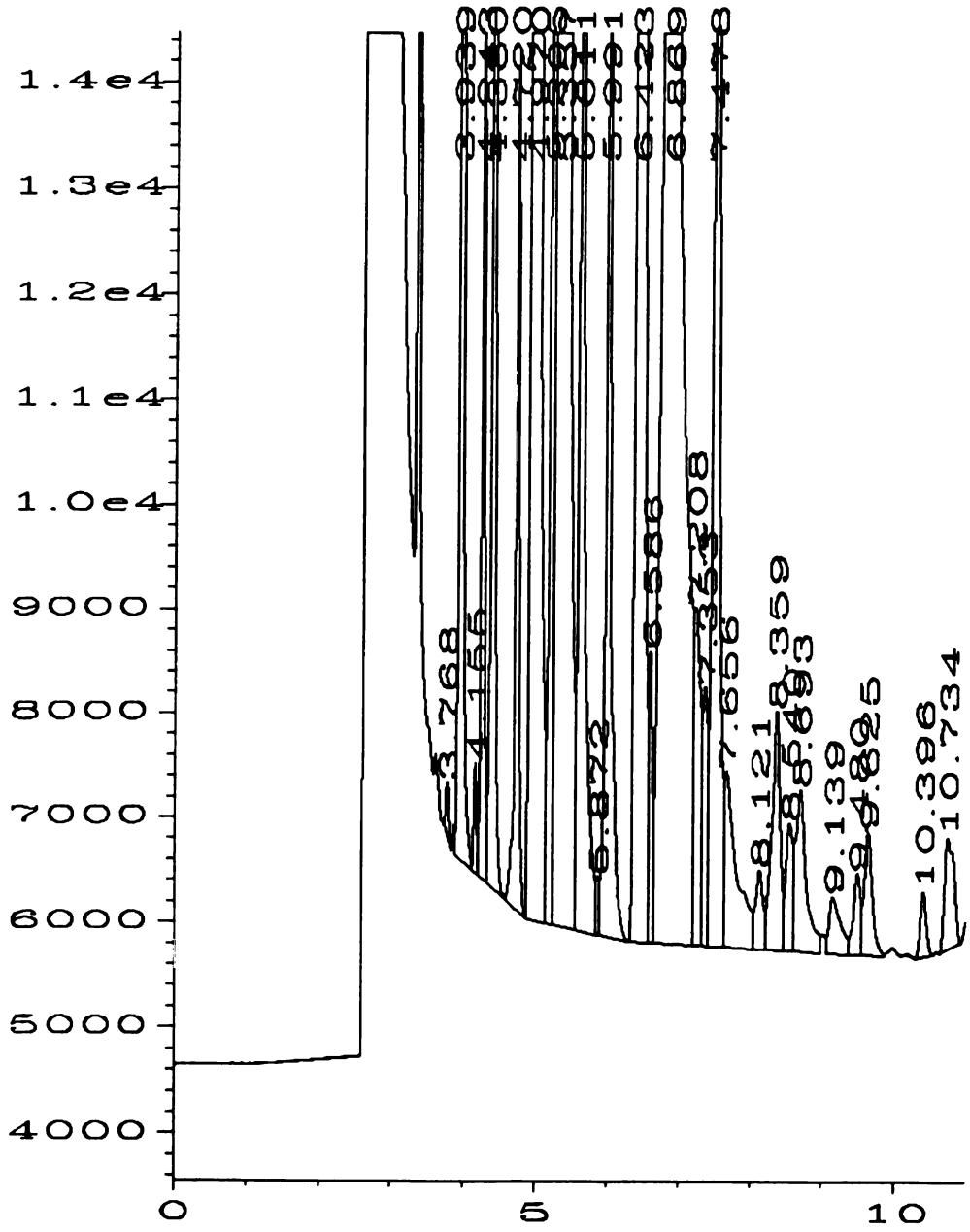


Figure 8. Gas chromatogram of fatty acid methyl esters from sample 143 (cultures 9A5 through 9A9). Note the elution of a small peak at 8.540 min, at the same elution time as authentic TMS-methyl ricinoleate (see Figure 7). This peak was not seen when cultures 9A5 through 9A9 were re-grown and analysed individually. (Attenuation and axes as for Figure 6).

Experiment 2

Clones in 96-well plates were replicated onto galactose-containing agar plates using a 96-prong device. The clones were grown to patches and then the cells from each plate were scraped together in 5 ml water. The cells from five plates (representing five 96-well plates) were used to inoculate a 400 ml culture of galactose-containing medium and grown at 30°C for approximately 24 h. The cells were pelleted, stored briefly at -20°C, and analysed for FAME profile by gas chromatography. Between 2500 and 3000 clones were grown and analysed by this method. No difference in the gas chromatographs could be detected before and after TMS-derivatization, indicating that no hydroxylated fatty acids were present. This indicates that ricinoleic acid did not constitute as much as 10% of the fatty acids of any individual culture.

DISCUSSION AND CONCLUSIONS

In experiments 1 and 2, a total of approximately 5000 yeast clones were analysed for fatty acid content. In experiment 1, 2300 clones showed no differences in the overall fatty acid profile, and no ricinoleic acid could be detected. The sensitivity and resolution of the gas chromatograph system was such that ricinoleic acid could have been detected if it were present at 0.1-0.5% of total fatty acids of an individual culture. In experiment 2, 2500 to 3000 clones were screened by a more rapid technique, where larger pools of cultures were first enriched for polar fatty

acids by chromatography on silicic acid. A reconstruction experiment indicated that ricinoleic acid would be detectable if it constituted 10% of the fatty acids of an individual clone. No ricinoleic acid was detected.

As explained in the introduction, this experiment was viewed as risky. It was intended only to screen a reasonable number of clones, on the assumption that clones for the hydroxylase should be moderately abundant, and to terminate the experiment if it appeared likely that some other problem precluded success. By the design of the experiment, the negative result obtained also does not provide much new information regarding the nature of the oleate-12-hydroxylase. The critical factors of the experimental design are discussed briefly to put the results in context.

Heterologous expression of a range of eukaryotic proteins was described in the introduction. In many cases, proteins are targetted in the yeast cell in a very similar manner to that in the native cell. Since the experiments reported here were completed, many more such reports of successful heterologous expression have appeared. These include plant soluble proteins[20-23] and correct targetting of plant mitochondrial proteins,[24,25] nuclear proteins,[26,27] and plasmamembrane proteins.[28-30] Furthermore, a plant cytochrome P450 has been functionally expressed in yeast, and is localised in microsomes where it forms an active complex with yeast NADPH:P450 reductase.[31] This suggests that the failure to detect functional expression of an oleate-12-hydroxylase in the experiments reported here is probably not due to a problem of heterologous expression.

Various other criteria for successful expression of an oleate-12-hydroxylase in

yeast were discussed in the introduction. It must be assumed that for one or more of these reasons, no expression could be detected. In addition to the hydroxylase, it was hoped that differences in fatty acid composition might also be detected due to the expression of another cDNA, such as a microsomal desaturase. However, this was also not observed. Difficulties have independently been encountered, however, with functional expression of a plant microsomal fatty acid desaturase in yeast (V.A. Arondel, unpublished). A putative oleate-12-hydroxylase clone also gave no functional expression in yeast (see chapter 6).

REFERENCES

1. Imai, Y., Characterization of rabbit liver cytochrome P-450 (laurate ω-1 hydroxylase) synthesized in transformed yeast cells, *J. Biochem. (Japan), 103, 143-148, 1988.*

2. Imai, Y., and Nakamura, M., The importance of threonine-301 from cytochromes P-450 (laurate (ω-1)-hydroxylase and testosterone 16α-hydroxylase) in substrate binding as demonstrated by site-directed mutagenesis, *FEBS Lett.*, 234, 313-315, 1988.

3. Imai, Y., Uno, T., Nakamura, M., and Yokota, H., Structure-function relationships of cytochrome P-450 laurate (ω-1)-hydroxylase, *Drug Metabolism Reviews*, 20, 467-478, 1989.

4. Uno, T., and Imai, Y., Identification of regions functioning in substrate interaction of rabbit liver cytochrome P-450 (laurate (ω-1)-hydroxylase), *J. Biochem. (Japan)*, 106, 569-574, 1989.

5. Oeda, K., Sakaki, T., and Ohkawa, H., Expression of rat liver cytochrome P-450MC cDNA in *Saccharomyces cerevisiae*, *DNA*, 4, 203-210, 1985.

6. Cullin, C., and Pompon, D., Synthesis of functional mouse cytochromes P-450 P1 and chimeric P-450 P3-1 in the yeast *Saccharomyces cerevisiae*, *Gene*, 65, 203-217, 1988.

7. Yellen, G., and Migeon, J.C., Expression of *Torpedo* nicotinic acetylcholine receptor subunits in yeast is enhanced by use of yeast signal sequences, *Gene*, 86,145-152, 1990.

8. Sogaard, S., and Svensson, B., Expression of cDNAs encoding barley α-amylase 1 and 2 in yeast and characterization of the secreted proteins, *Gene*, 94, 173-179, 1990.

9. Tague, B.W., Dickinson, C.D., and Chrispeels, M.J., A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole, *Plant Cell*, 2, 533-546, 1990.

10. Stukey, J.E., McDonough, V.M., and Martin, C.E., The *OLE1* gene of *Saccharomyces cerevisiae* encodes the $\Delta 9$ fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene, *J. Biol. Chem.*, 265, 20144-20149, 1990.

11. Zinser, E., Sperka-Gottlieb, C.D.M., Fasch, E.V., Kohlwein, S.D., Paltauf, F., and Daum, G., Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote *Saccharomyces cerevisiae*, *J. Bacteriol.*, 173, 2026-2034, 1991.

12. Rao, S.V., Paulose, M.M., and Vijayalakshmi, B., A note on the fatty acids present in oilseed phospholipids, *Lipids*, 2, 88, 1967.

13. Kondoh, H., and Kawabe, S., Phospholipids in castor seeds, *Agr. Biol. Chem.*, 39, 745, 1975.

14. Light, R.J., Lennarz, W.J., and Bloch, K., The metabolism of hydroxystearic acids in yeast, J. Biol. Chem., *237, 1793-1800, 1962.*

15. Shanklin, J. and Somerville, C., Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs, *Proc. Natl. Acad. Sci. USA*, 88, 2510, 1991.

16. Greenwood, J.S., and Bewley, J.D., Seed development in *Ricinus communis* (castor bean). I. Descriptive morphology, *Can. J. Bot.*, 60, 1751-1760, 1982.

17. Puissant, C., and Houdebine, L., An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *BioTechniques*, 8, 148-149, 1990.

18. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

19. Becker, D.M., and Guarente, L., High efficiency transformation of yeast by electroporation, *Meth. Enzymol.*, 194, 182-187, 1991.

20. Alderson, A., Sabelli, P.A., Dickinson, J.R., Cole, D., Richardson, M., Kreis, M., Shewry, P.R., and Halford, N.G., Complementation of *snf1*, a mutation affecting global regulation of carbon metabolism in yeast, by a plant protein kinase cDNA, *Proc. Natl. Acad. Sci. USA*, 88, 8602-8605, 1991.

21. Deforce, L., Tomizawa, K.-I., Ito, N., Farrens, D., Song, P.-S., and Furuya, M., *In vitro* assembly of apophytochrome and apophytochrome deletion mutants expressed in yeast with phycocyanobilin, *Proc. Natl. Acad. Sci. USA*, 88, 10392-10396, 1991.

22. Wahleithner, J.A., Li, L., and Lagarias, J.C., Expression and assembly of spectrally active recombinant holophytochrome, *Proc. Natl. Acad. Sci. USA*, 88, 10387-10391, 1991.

23. Blechl, A.E., Thrasher, K.S., Vensel, W.H., and Creene, F.C., Purification and characterization of wheat α-gliadin synthesized in the yeast, *Saccharomyces cerevisiae*, *Gene*, 116, 119-127, 1992.

24. Chaumont, F., O'Riordan, V., and Boutry, M., Protein transport into mitochondria is conserved between plant and yeast species, *J. Biol. Chem.*, 265, 16856-16862, 1990.

25. Huang, J., Struck, F., Matzinger, D.F., and Levings, C.S., Functional analysis in yeast of cDNA coding for the mitochondrial Rieske iron-sulfur protein of higher plants, *Proc. Natl. Acad. Sci. USA*, 88, 10716-10720, 1991.

26. Hirt, H., Pay, A., Gyorgyey, J., Bako, L., Nemeth, K., Bogre, L., Schweyen, R.J., Heberle-Bors, E., and Dudits, D., Complementation of a yeast cell cycle mutatn by an alfalfa cDNA encoding a protein kinase homologous to $p34^{cdc2}$, *Proc. Natl. Acad. Sci. USA*, 88, 1636-1640, 1991.

27. Tinland, B., Koukolikova-Nicola, Z., Hall, M.N., and Hohn, B., The T-DNA-linked VirD2 protein contains two didtinct functional nuclear localization signals, *Proc. Natl. Acad. Sci. USA*, 89, 7442-7446, 1992.

28. Anderson, J.A., Huprikar, S.S., Kochian, L.V., Lucas, W.J., Gaber, R.F., Functional expression of a probable *Arabidopsis thaliana* potassium channel in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. USA*, 89, 3736-3740, 1992.

29. Sentenac, H., Bonneaud, N., Minet, M., Lacroute, F., Salmon, J.-M., Gaymard, F., and Grignon, C., Cloning and expression in yeast of a plant

potassium ion transport system, *Science*, 256, 663-665, 1992.

30. Villalba, J.M., Palmgren, M.G., Berberian, G.E., Ferguson, C., and Serrano, R., Functional expression of plant plasma membrane H^+-ATPase in yeast endoplasmic reticulum, *J. Biol. Chem.*, 267, 12341-12349, 1992.

31. Bozak, K.R., O'Keefe, D.P., and Christoffersen, R.E., Expression of a ripening-related avocado (*Persea americana*) cytochrome P450 in yeast, *plant Physiol.*, 100, 1976-1981, 1992.

CHAPTER 4

EXPERIMENTS USING DESATURASE GENES IN ATTEMPTS TO ISOLATE THE HYDROXYLASE GENE

ABSTRACT

It is hypothesized that oleate-12-hydroxylase is homologous to the microsomal fatty acid desaturases, based upon three arguments. These are (a), that ricinoleic acid is found in isolated taxa of the plant kingdom, suggesting that oleate-12-hydroxylase has evolved recently and rapidly; (b), that biochemical studies identify similarities between the hydroxylase and desaturases; and (c), that the active site of the membrane-bound desaturases may contain a μ-oxo bridged diiron cluster, capable of catalysing hydroxylation as well as desaturation.

On this hypothesis, two approaches were used in attempts to clone the hydroxylase gene using desaturase genes. The *Brassica napus fad3* gene was used to screen a castor developing seed cDNA library at low stringency. One class of clones was isolated, encoding the castor *fad7* desaturase, for which a complete sequence was obtained.

Polymerase chain reaction was used to amplify sequences from castor developing seed mRNA which encode a motif conserved among desaturase proteins (GHDCGH). *Fad7* was the only desaturase or desaturase-like sequence isolated by this approach.

It is concluded that the oleate-12-hydroxylase gene, if homologous to the desaturase genes, is too divergent from them to be isolated using direct screening and PCR approaches.

INTRODUCTION

A feature of the biochemistry of fatty acid modifying enzymes involved in the biosynthesis of usual fatty acids (such as the desaturases) and of unusual fatty acids (such as castor oleate-12-hydroxylase) is that they catalyse reactions in which an unactivated C-H bond is cleaved. This cleavage is energetically demanding, and these fatty acid modifying enzymes utilise the high oxidising power of molecular oxygen. There are presently two known classes of enzyme cofactors capable of this type of O_2-dependent chemistry. The haem-containing oxygenases including cytochromes P450 are one class. These enzymes have been extensively characterised in animal and bacterial systems.

The mechanism of a hydroxylation reaction by a cytochrome P450[1] might be summarised as follows. Two electrons must be transferred to the cytochrome P450 to effect the reaction, but this occurs as two discrete single electron transfers. Binding of substrates to the resting state iron(III) enzyme increases the oxidation potential of the cofactor, potentiating the first electron transfer, leading to the iron(II) state. Oxygen binds to the iron(II) state, giving rise to an oxygenated intermediate described as a coordinated superoxide, Fe(III)-O_2^-. Alternatively, carbon monoxide can bind to

the iron(II) state, leading to the diagnostic inhibition of P450 enzymes. The second electron transfer reduces the superoxide form of the cofactor to a peroxide form, Fe(II)-OO^{2-}. Catalysis is now initiated, the peroxide form undergoing spontaneous heterolytic cleavage to generate water and the reactive intermediate Fe(V)=O. It is thought that this reactive intermediate is stabilised by an oxidisable ligand provided by the porphyrin ring or a tryptophan residue. Such stabilisation would be essential for the specificty of the reaction; in its absence the highly reactive intermediate would likely oxidise the nearest available atom in an uncontrolled fashion. Thus stabilisation of the Fe(V)=O intermediate allows the abstraction of a specific hydrogen atom from the substrate followed by a recombination that results in hydroxylation of the substrate and completion of the reaction cycle.

Research leading to this understanding of the mechanism of cytochrome P450 catalysis has provided an appreciation of the thermodynamic stability of unactivated aliphatic C-H bonds and the highly reactive oxoiron intermediates necessary for their cleavage. Non-haem cofactors performing similar chemistry will require similarly reactive intermediates which will also require stabilisation for a controlled reaction. In addition, the common intermediate used by different P450 enzymes indicates that reaction specificity must reside in the electronic and structural properties of the active site of the protein. In other words, the differing protein environments of different enzymes using the same cofactor results in their specific biochemistry.

We now consider the second class of cofactor known to be capable of this type of O_2-dependent chemistry. This class is less well characterised, and is typified by

the bacterial enzyme methane monooxygenase.[1] The cofactor in the hydroxylase component of methane monooxygenase is termed a μ-oxo bridged diiron cluster (FeOFe). The two iron atoms of the FeOFe cluster are liganded by protein-derived nitrogen or oxygen atoms, and are tightly redox-coupled by the covalently-bridging oxygen atom. The catalytic cycle of methane monooxygenase is not so well understood as that of the P450 oxygenases, but there are known differences and similarities. Rather than two discrete single-elctron reductions of the haem cofactor, the FeOFe cluster accepts two electrons, reducing it to the diferrous state, before oxygen binding. Upon oxygen binding, it is likely that heterolytic cleavage also occurs, leading to a high valent oxoiron reactive species that is very smilar to that of the haem cofactor, but stabilised by resonance rearrangements possible within the tightly coupled FeOFe cluster, rather than through a porphyrin- or protein-derived ligand. The stabilised high-valent oxoiron state of methane monooxygenase is capable of proton extraction from methane, followed by oxygen transfer, giving methanol.

The FeOFe cofactor has been shown to be directly relevant to plant fatty acid modifications by the demonstration that castor stearoyl-ACP desaturase contains this type of cofactor.[2] This desaturase is unique among the usual plant fatty acid desaturases in being a soluble enzyme, whereas the other usual desaturases are membrane-bound. The protein was purified from avocado fruit and a cDNA was isolated from developing castor endosperm.[3] Putative iron-binding motifs have been identified in the stearoyl-ACP desaturase primary structure by comparison to other soluble enzymes containing the FeOFe cluster.[2] These similar motifs, (D/E)-E-X-R-

H, are characteristically spaced approximately 90 residues apart in a number of soluble diiron-oxo proteins, including methane monooxygenase. Since the work described in chapter 3 was done, the first cDNA clone for a plant membrane-bound desaturase was isolated in our laboratory,[4] encoding the microsomal ω-3 desaturase of *Brassica napus*. Of great interest is the identification of a similarly repeated motif in both this sequence, the membrane-bound rat stearoyl-CoA desaturase,[5] and in two membrane-bound monooxygenases.[6,7] This motif, H-X-X-H-H in the desaturases and H-X-X-X-H-H in the monooxygenases, may be the functional equivalent in membrane-bound FeOFe proteins of the (D/E)-E-X-R-H motif in the soluble FeOFe proteins. Evidence in support of this view is that replacement with asparagine of any of the histidine residues of these motifs in the rat stearoyl-CoA desaturase, abolishes enzyme activity (J. Shanklin, personal communication). This suggests that the plant membrane bound desaturases may also accomplish oxygen-dependent fatty acid desaturation through an FeOFe cofactor. All membrane-bound plant desaturases subsequently sequenced also contain the conserved histidine-rich repeats (below, and references 8,9).

Of the well-characterised FeOFe-containing enzymes, methane monooxygenase catalyses a reaction involving oxygen-atom transfer ($CH_4 \rightarrow CH_3OH$), while the FeOFe cluster of ribonucleotide reductase catalyses the oxidation of tyrosine to form a tyrosyl cation radical without oxygen-atom transfer. However, site-directed mutagenesis of Phe208 to Tyr resulted in the conversion of this enzyme to an oxygen transfer catalyst, Tyr208 being hydroxylated and shown to be acting as a ligand to

one iron of the FeOFe cluster.

The argument made for the P450 oxygenases catalysing a range of reactions through the use of the same reactive intermediate modulated by the electronic and structural environment provided by the protein, might also be applied to FeOFe-containing enzymes. Modifications of the active site of plant fatty acid oxidising enzymes containing FeOFe clusters could thus alter the outcome of the reaction, including whether oxygen-atom transfer occurs or not.

Hypothesis For the Origin of Oleate-12-Hydroxylase

An hypothesis is now proposed for the origin of castor oleate-12-hydroxylase, based upon three arguments. The first argument involves the taxonomic distribution of plants containing ricinoleic acid, mentioned in chapter 1. Ricinoleic acid has been found in 12 genera of 10 families of higher plants.[10-21] Thus, plants in which ricinoleic acid occurs are found throughout the plant kingdom, yet close relatives of these plants do not contain the unusual fatty acid. This pattern suggests that the ability to synthesize ricinoleic acid has arisen several times independently, and is therefore a quite recent divergence. In other words, the ability to synthesize ricinoleic acid has evolved rapidly, suggesting that a relatively minor genetic change was necessary to accomplish it. Two mechanisms for such facile evolution of a new enzyme activity are envisaged. One mechanism would be for the modification of a gene normally encoding a fatty acid hydroxylase active in the epidermis and involved in the synthesis of a hydroxy-fatty acid cutin monomer. The other mechanism would

be for modification of a gene encoding a microsomal fatty acid desaturase, such that instead of performing one type of oxidation reaction (desaturation) it now performs another (hydroxylation).

The second argument is that many biochemical properties of castor oleate-12-hydroxylase are similar to those of the microsomal desaturases, as discussed in chapter 2. Furthermore, biochemical studies of the hydroxylase also suggest that it is not a P450 enzyme. The fatty acid hydroxylases known to be involved in synthesis of cutin monomers are cytochromes P450.[22]

The third argument stems from the discussion of oxygenase cofactors above, in which it is suggested that the plant membrane bound fatty acid desaturases may have a μ-oxo bridged diiron cluster-type cofactor, and that such cofactors are capable of catalysing both fatty acid desaturations and hydroxylations, depending upon the electronic and structural properties of the protein active site.

Taking these three arguments together, it is now hypothesized that oleate-12-hydroxylase of castor endosperm is homologous to the microsomal oleate $\Delta 12$ desaturase found in all plants. This is in accord with the apparently rapid evolution of the hydroxylase gene. However, it should be noted that this hypothesis is proposed specifically for the castor enzyme, for which biochemical studies have shown the similarity of many properties of the hydroxylase and desaturases. Other species accumulating ricinoleic acid, for which no biochemical studies are available, may have evolved an hydroxylase by a different route, such as by modification of an existing cytochrome P450 gene.

The above hypothesis proposes that the oleate-12-hydroxylase gene is homologous to the gene of the biochemically most-similar desaturase, the microsomal Δ12 desaturase. However, it was thought at this time that this gene, in turn, was probably homologous to the other microsomal (ω-3) desaturase in particular, and possibly also to the equivalent desaturases of the chloroplast inner envelope. A gene (*fad3*) had recently been isolated in our laboratory for the microsomal ω-3 desaturase from *Brassica napus*,[4] hence I sought to use this clone in various attempts to isolate homologous genes from castor, one of which might encode the oleate-12-hydroxylase. These experiments are the subject of this chapter. First, a castor genomic Southern blot is probed with the *fad3* clone to confirm that a number of homologous sequences exist in castor. Attempts are then made to isolate these by direct screening and polymerase chain reaction (PCR) approaches.

MATERIALS AND METHODS

Southern Analysis of Castor Genomic DNA Using the *fad3* Probe

The *Brassica napus fad3* clone was labelled by random priming[23] and purified of unincorporated nucleotides by ethanol precipitation in the presence of ammonium acetate.[23] A Southern blot of digested castor genomic DNA was generously provided by Dr. J. Shanklin[3]. The blot was prehybridised at 52°C in a solution containing 4 X SET (0.6 M NaCl, 0.12 M Tris-HCl pH 7.4, 8 mM EDTA), 0.1% sodium

pyrophosphate, 0.2% SDS, and 100 μg ml^{-1} heparin.[24] The probe was hybridised to the blot at 52°C overnight in the same solution, except for the addition of 10% dextran sulphate. The blot was washed three times in 2 X SSC, 0.1% SDS at 52°C, then exposed to X-ray film.

Screening of cDNA Library with *fad3* Probe

Five plates of each of the four pools of the pYES2.0 cDNA library (chapter 3) were grown until small colonies ($\sim 10^5$ total) were visible. A nitrocellulose filter (Schleicher & Schüll BA85) was laid on each plate, its position marked, and lifted off to a fresh plate, the adhering colonies now facing upwards. Care was necessary that both plate and filter were not too moist, to avoid smearing of the colonies. The original plate was incubated for a few h at 37°C to recover colonies, while the filters were processed as folllows. Each filter was sequentially placed, colony side up, on Whatman 3MM paper moist with 10% SDS (3 min), denaturing solution (0.5 M NaOH, 1.5 M NaCl; 5 min), neutralising solution (0.5 M Tris-Cl pH 7.4, 1.5 M NaCl; 5 min), and 2 X SSC (0.3 M NaCl, 0.03 M Na-Citrate, pH 7.0). The filters were then air-dried for ca. 1 min before pressing twice between sheets of filter paper to remove cell debris. After air-drying a further 30 min, DNA was fixed to the filters by baking *in vacuo* at 80°C for 1-2 h.

A probe was prepared in the same manner as for the Southern blot described above, and hybridised to the filters overnight at 55°C in a solution with the same

composition as the prehybridisation solution described above for the Southern blot. The filters were washed at 55°C in the solution described above and exposed to X-ray film. Hybridizing colonies were picked and replated for a second round of hybridization, yielding positive clones. Minipreparations of DNA from these clones was simultaneously digested with *Bam*H1, *Eco*R1, and *Xho*1, releasing and possibly fragmenting the insert of each clone, and electrophoresed through agarose gels. A Southern blot was prepared from these gels as follows. DNA in the gels was denatured by incubation for ~1 h in 0.5 M NaOH, 1.5 M NaCl, and then the gels were neutralised by two 30 min incubations in 0.5 M Tris-Cl pH 7, 1.5 M NaCl. DNA from the gels was transferred overnight onto a nylon membrane (Hybond N, Amersham) by elution from the gel in 10 X SSC (1.5 M NaCl, 0.15 M Na-citrate, pH 7.0). After air-drying for 1 h, DNA was fixed to the membrane by exposing to a UV light source for 2 min (based upon an empirical calibration). The *fad3* probe was prepared and hybridised (at 55°C) to the Southern blot as described above.

Generation of Nested Deletions and DNA Sequencing

Plasmid DNA was prepared from large (500 ml) cultures of *E. coli* harbouring pFL1 or pFL1r by the alkaline lysis technique and purified by banding in CsCl density gradients as described.[23] Nested sets of deletions were generated from these clones by the exonuclease III protocol.[23] The enzymes used to digest the plasmid DNA were *Bam*H1 and *Kpn*1. Deletion clones varying in size by approximately 250

bp were sequenced. Plasmid DNA for sequencing was prepared from overnight cultures (5 ml, LB medium containing 100 mg l^{-1} ampicillin) using "Magic Minipreps" (Promega) according to the instructions of the manufacturer, and submitted for automated cycle sequencing on Applied Biosystems 373A instruments using the fluorescent primer T7. Sequence data was aligned, assembled and analysed with the DNASIS and PROSIS computer programs.

Reverse Transcription-Polymerase Chain Reaction

Poly(A)$^+$ RNA from developing castor endosperm was that used for construction of the pYES2.0 cDNA library (chapter 3). Leaf poly(A)$^+$ RNA was purified by the same technique (chapter 3) from small true leaves of seedling plants, from which the midribs were removed.

Oligonucleotide primers used in these experiments were as follows:

R_1T_{15}: 5' GAC ATC GAT AAT ACT TTT TTT TTT TTT TT 3'

R_1: 5' GAC ATC GAT AAT AC 3'

GO: 5' GGN CA(C/T) GA(C/T) TG(C/T) GGN CA 3'.

Reverse transcription reactions were as follows. RNA (2 μg, 10.25 μl) was heated to 65°C for 3 min, then cooled on ice, and added to a tube containing the other reaction components:

4 μl 5X reverse transcriptase buffer (BRL)

2 μl 100 mM dithiothreitol

0.25 μl (~ 10 units) RNAsin (Promega)

0.5 μl (2.5 pmol) R_1T_{15} primer

2 μl 10 mM each dNTP.

The reaction was initiated by addition of 1 μl (200 units) reverse transcriptase (BRL, from murine Maloney leukemia virus) and incubated at 37°C for 2 h. The reaction was then diluted with 980 μl TE and stored at 4°C.

PCR reactions contained 250 ng each primer (R_1, GO), 0.2 mM each dNTP, 7 mM $MgCl_2$, Taq polymerase buffer (Promega). Taq polymerase (2.5 units, Promega) and first-strand cDNA template (5 μl) were added, bringing total volume to 50 μl, at 75°C to prevent mis-priming at low temperature. Cycling was essentially according to Frohman *et al*[25]: first cycle: 95°C for 5 min, 50°C 2 min, 72°C 40 min; second cycle: 95°C 45s, 50°C 25 s, 72°C 3 min (with addition of 2.5 units additional Taq polymerase); followed by twenty-nine identical cycles (no further enzyme addition) and a final 15 min extension at 72°C.

RESULTS AND DISCUSSION

Southern Analysis of Castor Genomic DNA Using the *fad3* Probe

A genomic castor Southern blot was probed with the *fad3* clone from *Brassica napus*, at moderately low stringency (52°C). Approximately 9 hybridizing bands could be detected in an *Eco*R1 digest (Figure 9), indicating that at least several genes in the castor genome share homology with the *fad3* probe, as expected.

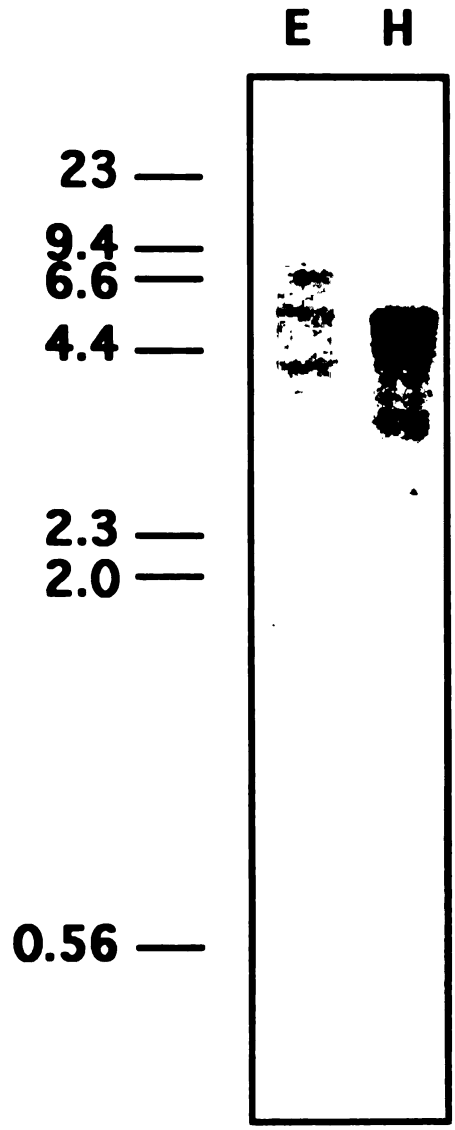


Figure 9. A Southern blot of castor genomic DNA digested with restriction enzymes *Eco*RI (E) or *Hin*dIII (H), was hybridised at moderately low stringency (52°C, 4 X SET) with the ^{32}P-labelled *Brassica napus fad3* cDNA. Migration of DNA standards (kb) is shown to the left.

Screening of cDNA Library with *fad3* Probe

Approximately 10^5 colonies of the pYES2.0 cDNA library, constructed in chapter 3, were screened with the *Brassica napus fad3* clone. The hybridization temperature was slightly higher (55°C) than that used for the Southern blot (52°C). Low stringency screening of *E. coli* colonies was frequently problematic, binding of the probe to bacterial debris remaining on the filters frequently leading to unacceptably high background levels. Twenty-five apparent positives were re-screened, giving seven apparently true positives. However, when the plasmid DNA of these clones was digested and analysed by Southern blot, the *fad3* probe hybridised specifically to the insert DNA of only three clones. These three clones gave the same restriction pattern but varied in total length. The longest clone, isolate 3ei,b, was approximately 2.0 kb, designated pFL1, and selected for full-length sequencing. The insert of pFL1 was excised with *Eag*1, which cuts on both sides of the polylinker, and religated. A clone in which the insert was now in the opposite orientation to pFL1 was selected and designated pFL1r. A deletion series was generated from both pFL1 and pFL1r, such that by sequencing with the T7 primer the complete sequence of both strands was obtained. The clone is 1958 nucleotides in length and consists of a 344 nucleotide 5' untranslated region, a 1380 nucleotide open reading frame, and a 234 nucleotide 3'untranslated region. The possibility that the unusually large 5' untranslated region is due to some cloning artifact was not investigated. The open reading frame encodes a 460 amino acid protein with a calculated M_r of 52558.

While this work was in progress, additional desaturase clones were being isolated by other members of the laboratory, including the *fad7* gene of *Arabidopsis thaliana*. This gene encodes a plastid ω-3 desaturase.[9] The gene of a membrane-bound Δ12 desaturase (*desA*) from the cyanobacterium *Synechococcus* was also known.[26] Comparison of the pFL1 sequence with these other sequences is shown schematically in Figure 10 (see also Figure 11). At the nucleotide level, pFL1 shares the highest sequence similarity to a central region of the *fad7* sequence, and respectively lower similarities to the *fad3* and *desA* sequences in approximately the same region. There is an additional region of moderate similarity between the pFL1 sequence and the 5' end of the *fad7* sequence, not found for *fad3* or *desA*. These results are obtained also by comparison at the amino acid level, similarity being highest to *fad7* and also extending further toward the amino terminus than for either the *fad3* or *desA* sequences. Thus, the amino-terminus of the predicted castor protein has an extension similar to that of the *fad7* protein but not found in the *fad3* protein. This amino-terminal extension is rich in the hydroxy amino acids serine and threonine (23% of the first 78 residues), a characteristic feature of the transit peptide of plastid proteins.[27] These results suggest that pFL1 encodes a plastid ω-3 desaturase homologous to *fad7*. This sequence has been deposited in Genbank (accession L25897), and a brief report submitted for publication.[28] Other ω-3 desaturase genes have been reported recently.[8]

The sequence similarity between the ω-3 desaturases demonstrates the high degree of conservation between the endoplasmic reticulum and plastid forms. This

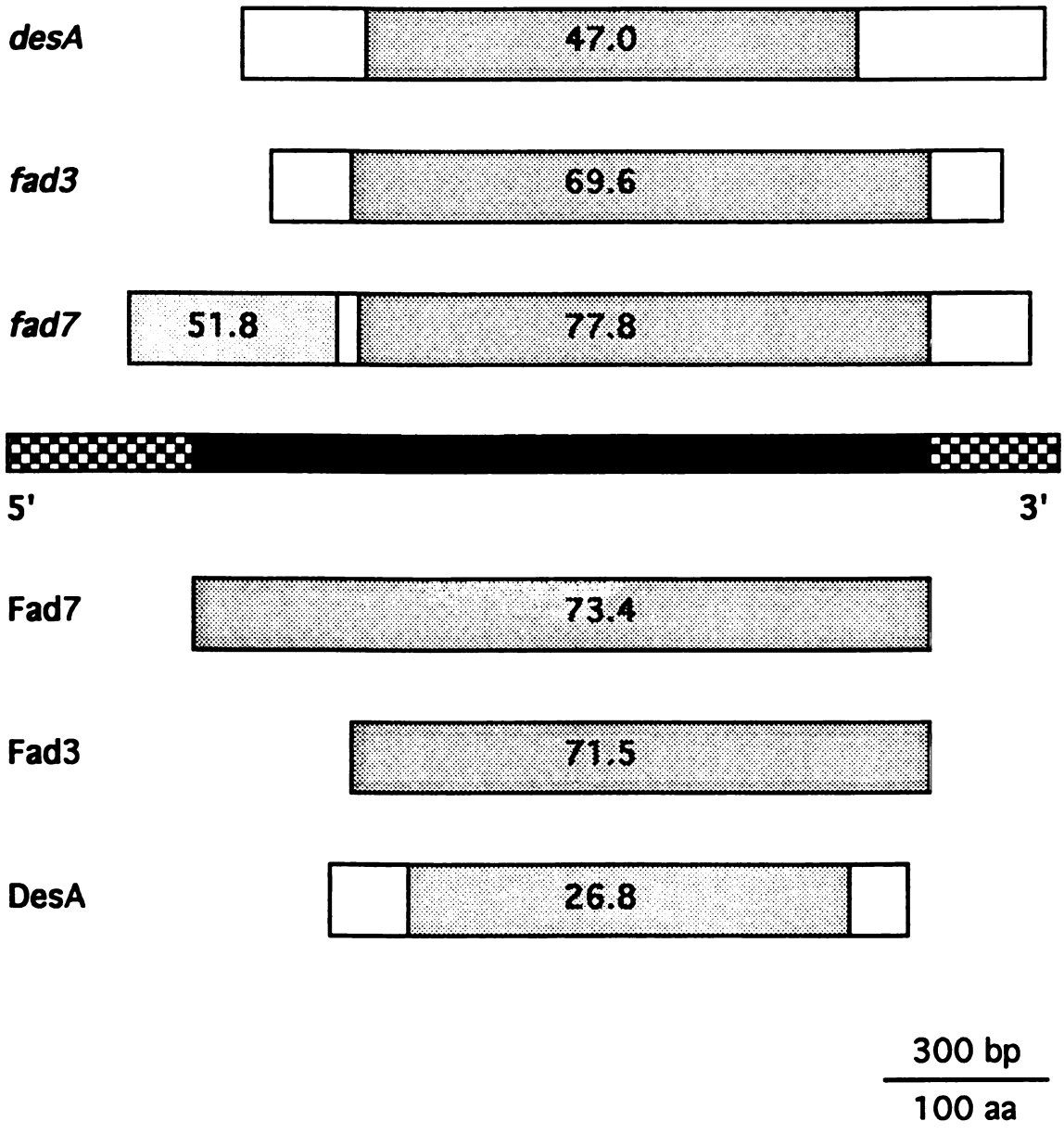


Figure 10. Sequence similarity between pFL1 (centre) and other membrane-bound desaturases at the nucleotide (above) and amino acid (below) levels. The pFL1 coding sequence and untranslated regions are shown as solid and checked bars, respectively. The percent sequence similarity between pFL1 and the other genes is shown in the shaded regions over which the similarity was averaged.

was not the expected result when this experiment was embarked upon. The same result has also been obtained by other members of the laboratory, that is, that desaturases acting at the ω-3 position are conserved among themselves, but do not serve as heterologous probes for those acting at the Δ12 position, even when probing within a species (*Arabidopsis*) (S. Gibson, personal communication). The evolution of the desaturases, therefore, appears to have involved an earlier differentiation between the Δ12 and ω-3 enzymes, and a subsequent differentiation for each type giving rise to the microsomal and plastid forms. Since it is hypothesized that the castor oleate-12-hydroxylase gene is homologous to the microsomal Δ12 desaturase, only a desaturase probe capable of isolating the microsomal Δ12 desaturase gene should also be a possible probe for the hydroxylase. Our experiments show that *fad3* is not such a probe. In addition, one screening experiment was done with the *Arabidopsis* clone GO313. This desaturase homologue was isolated by Dr. S. Gibson in our laboratory, and is tentatively assigned as the gene (*fad6*) for the plastid Δ12 desaturase. Of 80 000 lambda phage clones screened from the λZAPST library (described in chapter 5) at 52°C, no duplicate positives were obtained. These results indicate that an alternative strategy is required for cloning the hydroxylase on the basis of possible homology to desaturases.

```
                    10         20         30         40         50
RCFAD7        1 MAAGWVLSEC GLRPLPRIYS RPRIGFTSKT TNLLKLRELP DSKSYNLCSS   50
ATFAD7       -1 .M*NL***** *I********T T**SN*L*NN NKFRPSLSSS SY*TSSSPL*   49
BNFAD3      -79 .......... .......... .......... .......... ..........  -30
SDESA       -86 .......... .......... .......... .......... ..........  -37

                    60         70         80         90        100
RCFAD7       51 FKVSSWSNSK QSNWALNVAV PVNVSTVSGE DDREREEFNG IVNVDEGKGE  100
ATFAD7       50 *GLN*RDGFT R-*******ST *--------- --LTTPI*EE -SPLE*DNKQ   99
BNFAD3      -29 .......... .......... .........M VVAMDQRS*V NGDSGAR*E*   21
SDESA       -36 .......... .......... .......... ......MTAT *PPLTPTVTP   14

                   110        120        130        140        150
RCFAD7       101 FFDAGAPPPF TLADIRAAIP KHCWVKNPWR SMSYVLRDVV VVFGLAAVAA  150
ATFAD7       100 R**P****** N********* *********K *L***V***A I**A***G**  149
BNFAD3        22 G**PS*Q*** KIG******* ******S*L* *****T**IF A*AA**MA*V   71
SDESA         15 SNPDRPIADL K*Q**IKTL* *E*FE*KASK AWAS**ITLG AI-AVGYLGI   64

                   160        170        180        190        200
RCFAD7       151 YFNNWVAWPL YWFCQGTMFW ALFVLGHDCG HGSFSNNPKL NSVVGHLLHS  200
ATFAD7       150 *L***IV*** **LA****** ********** ******D*** **********  199
BNFAD3        72 **DS*FL*** **VA***L** *I******** *****DI*L* ******I***  121
SDESA         65 IYLP*YCL*I T*IWT**ALT GA**V***** *R**AKKRWV *DL***IAFA  114

                   210        220        230        240        250
RCFAD7       201 SILVPYHGWR ISHRTHHQNH GHVENDESWH PLSEKIFKSL DNVTKTLRFS  250
ATFAD7       200 ********** ********** ********** *M****YNT* *KP*RFF**T  249
BNFAD3       122 F********* ********** *********V **P**LY*N* PHS*RM**YT  171
SDESA        115 PLIY*F*S** LL*DH**LHT NKI*V*NA*D *W*VEA*QAS PAI---V*--  164

                   260        270        280        290        300
RCFAD7       251 LPFPMLAYPF YLWSRSPGKK GSHFHPDSGL FVPKERKDI- ITSTACWTAM  300
ATFAD7       250 **LV****** ***A****** ***Y****D* *L******V- L*********  299
BNFAD3       172 V*L******I ***Y******E ****N*Y*S* *A*S***L*- A***T**SI*  221
SDESA        165 *FYRAIRG** --*WTGSIFH W*LM*FKLSN *AQRD*NKVK LSIAVVFLFA  214

                   310        320        330        340        350
RCFAD7       301 AALLVYLNFS MGPVQMLKLY GIPYWIFVMW LDFVTYLHHH GHEDKLPWYR  350
ATFAD7       300 *****C***T I**I****** ******N*** ********** **********  349
BNFAD3       222 L*T****S*L VD**TV**V* *V**I***** **A******* **DE******  271
SDESA        215 *IAFPA*IIT T*VWGFV*FW LM*WLVYHF* MSTF*IV**T IP*IRF--RP  264

                   360        370        380        390        400
RCFAD7       351 GKAWSYLRGG L-TTLDRDYG -WINNIHHDI GTHVIHHLFP QIPHYHLVEA  400
ATFAD7       350 **E******* *-******** -L******** ********** **********  399
BNFAD3       272 **E******* *-**I***** -IF******* ********** ********D*  321
SDESA        265 AAD**AAEAQ *NG*VHC**P R*VEVLC*** NV*IP***SV A**S*N*RL*  314

                   410        420        430        440        450
RCFAD7       401 TEAAKPVMGK YYREPKKSGP LPLHLLGSLV RSMKEDHYVS DTGDVVYYQK  450
ATFAD7       400 *******L** *****D**** ******EI*A K*I******* *E*E****KA  449
BNFAD3       322 *R***H*L*R ******T**A I*I**VE*** A*I*K***** ****I*F*ET  371
SDESA        315 HGSL*ENW*P FLY*RTFNWQ *MQQIS*QCH LYDP*HG*RT FGSLKKV*..  364

                   460        470        480        490        500
RCFAD7       451 DPKLSGIGGE KTE..... .......... .......... ..........  500
ATFAD7       450 **N*Y*EVK* .......... .......... .......... ..........  499
BNFAD3       372 **D*YVYASD *SKIN..... .......... .......... ..........  421
SDESA        365 .......... .......... .......... .......... ..........  414
```

Figure 11. Comparison of the deduced amino acid sequences of four membrane bound desaturase genes: RCFAD7 *Ricinus communis* Fad7 (pFL1, above); ATFAD7 *Arabidopsis thaliana* Fad7;[9] BNFAD3 *Brassica napus* Fad3;[4] SDESA *Synechococcus* DesA.[26] The conserved GHDCGH and HXXHH motifs are shaded.

Amplification of Sequences With Conserved Desaturase Motifs by Polymerase Chain Reaction

A comparison of different desaturase genes (Figure 11) reveals a short stretch of perfectly conserved amino acids (GHDCGH). This sequence is found in the sameposition in the cyanobacterial Δ12 desaturase *desA* as well as the plant ω-3 desaturases *fad3* (from *Brassica napus*) and *fad7* (from both *Arabidopsis thaliana* and *Ricinus communis*). This suggested that this sequence may be directly involved in an aspect of the enzyme function (such as substrate binding), and might, therefore, be conserved among all desaturases and functional homologues, including the oleate-12-hydroxylase. An experiment was, therefore, designed to isolate sequences expressed in the developing castor endosperm that contain this conserved motif, using the polymerase chain reaction (PCR). Furthermore, ricinoleic acid, typical of unusual fatty acids (chapter 1), is specific to the seed tissue of castor, and is not found in vegetative tissues or synthesized in germinating seeds.[29-32] By comparing sequences amplified from both leaf and seed tissue, I sought to identify the hydroxylase in particular. Since it was known from the preceding experiment that the castor *fad7* homologue is expressed in the seed and contains this motif, a control for the experiment was that it should be possible at least to re-isolate this sequence.

Poly(A)$^+$ RNA purified from leaves and developing endosperm of castor was used as a template for reverse transcription. The reaction was primed with the oligonucleotide R_1T_{15}. A degenerate oligonucleotide ("Golden Oligo", GO) was

designed corresponding to all possible codons for the conserved motif GHDCGH. PCR reactions were then initiated using the first-strand cDNAs as templates and R_1 and GO as primers. The expected product size for an amplified desaturase fragment is 700-900 bp. Reaction conditions, particularly annealing temperature and magnesium concentration, were optimised to give a comprehensive range of products when analysed by agarose gel electrophoresis, and dependency of amplification upon both primers. Two particular difficulties were encountered. Gels of the PCR products were invariably smeared due to an unknown problem, making it difficult to identify individual bands. Secondly, where putative seed-specific bands were identified and isolated from the gel, they were completely resistant to cloning; all clones isolated from these ligations being of a different size than expected from the band size isolated from the gel, and invariably containing extraneous (bacterial or lambda phage) DNA. Hence an alternative approach was taken, simply cloning the seed-RNA PCR products directly, followed by analysis of the clones obtained. Difficulties were also encountered with these ligations and transformations, but a total of eleven clones apparently containing bona fide plant DNA could be isolated and were submitted for terminal sequencing, summarised in Table 6. The primer sequence used for 3' end amplification could be identified in eight of the eleven clones, and a partial primer sequence (matching 14 nucleotides of the 3' end of the 17-nucleotide primer) was identified in one further sequence. Since the PCR products were not directionally cloned, were sequenced only from one end, and in some cases appear to be too large to sequence through in a single sequencing run, it was not

Table 6. Identification (blastx database search result) of sequences cloned by PCR amplification using a sequence conserved among desaturases.

Clone	Sequence #	Highest Blast Score	
pIIIH6	CRS107	52	Non-structural polyprotein, rabbit hemorrhagic disease virus.
pIB9	CRS108	59	35 kDa major secreted protein precursor, vaccinia virus.
pIE1	CRS109	269	Stress-inducible protein STI35, *Fusarium solani*.
pIF7	CRS110	Same sequence as above (CRS109).	
pIH2	CRS112	77	ORF515 gene product, *Pinus contorta* chloroplast.
pIIC6	CRS114	367	60S ribosomal protein L5A, *Xenopus laevis*.
pIIF2	CRS116	Same sequence as CRS108, above.	
pIIIA4	CRS238	172	Linoleic acid desaturase, *Brassica napus*.
pIB5	CRS243	49	NADH-ubiquinone oxidoreductase chain 4, *Paramecium tetraurelia*.
p6-1	CRS247	61	Respiratory nitrate reductase 2 gamma chain, *Escherichia coli*.
p12-c	CRS254	59	DL-hydantoinase (hyuR) gene, *Pseudomonas* sp.

expected that the primer sequence would be identified in every case. The identification of the primer sequence in many of the clones suggests that castor cDNAs containing the primer sequence were indeed amplified.

DNA sequences were compared in all reading frames to the non-redundant translated-nucleotide and protein sequence databases (swissprot, PIR, GenBank, and EMBL) by the program blastx.[33] Blast scores lower than 80 are considered here to be insignificant, the corresponding sequences therefore having no matches in the databases. Most of the clones thus contain unidentified DNA. Three clones contained sequences similar to some other sequence in the databases. One of these, pIIIA4, was similar to the *fad3* desaturase of *Brassica napus*. In fact, this sequence is identical to a portion of the *fad7* sequence isolated from castor in the preceding experiment.

Thus, it is possible, as predicted in the introduction to this experiment, to isolate at least the *fad7* desaturase sequence from castor by this PCR approach using a conserved desaturase motif. However, no new desaturase or desaturase-like sequence was obtained in this experiment. The developing castor endosperm contains relatively small amounts of linolenic acid, the product of the *fad7* desaturase, hence high mRNA levels for this particular desaturase are not expected in this tissue. Indeed, in the preceding experiment, only three *fad7* clones were isolated from approximately 10^5 colonies of the cDNA library made from the same RNA. Based on the much higher activity of oleate-12-hydroxylase in the developing endosperm, and upon the hypothesis that hydroxylase and desaturase turnover numbers should be similar (see

chapter 3), it was expected that mRNA for oleate-12-hydroxylase should be more abundant than that for *fad7*. Thus the fact that only the desaturase was cloned by the PCR approach suggests that the conserved motif (GHDCGH) may not be highly conserved in the hydroxylase even if the hydroxylase is homologous to the desaturases.

Sequencing of other desaturase genes in our and other laboratories following the completion of this experiment supports this possibility. The *Arabidopsis* clone GO313 isolated by Dr. S. Gibson, tentatively identified as the plastid Δ12 desaturase, contains instead the divergent sequence GHDCAH, while an unidentified desaturase-homologue cloned from parsley in the laboratory of Dr. K. Hahlbrock, ELI72, contains instead the divergent sequence GHECDH (I. Sommsich, personal communication).

CONCLUSIONS

Attempts described here to isolate a clone for oleate-12-hydroxylase based upon hypothesized homology between this enzyme and microsomal desaturases, resulted instead in the isolation of a plastid desaturase gene. The sequence of this gene, *fad7*, demonstrated that genes encoding the same enzyme are strongly conserved between species (such as between castor and *Arabidopsis*). However, sequence similarity between different desaturases appears too weak in some cases to allow direct screening approaches. Thus it has not been possible to use the desaturase gene

fad3 as a probe for the isolation of the gene *fad2*. Likewise, in the experiments described here, it was not possible to isolate a clone for oleate-12-hydroxylase using *fad3* as a probe. On the basis of these results, it is not possible to distinguish whether oleate-12-hydroxylase is not homologous to the desaturases (contrary to the hypothesis), or whether they are homologous but considerably divergent. On the basis of the divergence observed among desaturases, the latter possibility is favoured.

REFERENCES

1. van de Loo, F.J., Fox, B.G., and Somerville, C., Unusual fatty acids, in *Lipid Metabolism in Plants,* T.S. Moore Jr., Ed., CRC Press, Boca Raton, 1993, chap 3.

2. Fox B.G., Shanklin, J., Somerville, C., and Munck, E., Stearoyl-acyl carrier protein Δ^9 desaturase from *Ricinus communis* is a diiron-oxo protein, *Proc. Natl. Acad. Sci.*, 90, 2486-2490, 1993.

3. Shanklin, J. and Somerville, C., Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs, *Proc. Natl. Acad. Sci. USA*, 88, 2510, 1991.

4. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M., and Somerville, C.R., Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*, *Science,* 258, 1353-1355, 1992.

5. Thiede, M.A., Ozols, J., and Strittmatter, P., Construction and sequence of cDNA for rat liver stearoyl coenzyme A desaturase, *J. Biol. Chem.*, 261, 13230-13235, 1986.

6. Kok, M., Oldenhuis, R., van der Linden, M.P.G., Raatjes, P., Kingma, J., van Lelyveld, P.H., and Witholt, B., The *Pseudomonas oleovorans* alkane hydroxylase gene: sequence and expression, *J. Biol. Chem.*, 264, 5435-5441, 1989.

7. Suzuki, M., Hayakawa, T., Shaw, J.P., Rekik, M., Harayama, S., Primary structure of xylene monooxygenase: similarities to and differences from the

alkane hydroxylation system, *J. Bacteriol.*, 173, 1690-1695, 1991.

8. Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Perez-Grau, L., Kinney, A.J., Hitz, W.D., Booth, R., Schweiger, B., Stecca, K.L., Allen, S.M., Blackwell, M., Reiter, R.S., Carlson, T.J., Russell, S.H., Feldmann, K.A., Pierce, J., and Browse, J., Cloning of higher plant ω3 fatty acid desaturases, *Plant Physiol.*, 103, 467-476, 1993.

9. Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S., and Somerville, C., A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana, J. Biol. Chem.*, in press.

10. Brieskorn, C.H., and Kabelitz, L., Hydroxyfettsauren aus dem cutin des blattes von Rosmarinus officinalis, *Phytochemistry*, 10, 3195, 1971.

11. Green, A.G., The occurrence of ricinoleic acid in Linum seed oils, *J. Am. Oil Chem. Soc.*, 61, 939, 1984.

12. Garg, S.P., Sherwani, M.R.K., Arora, A., Agarwal, R., and Ahmad, M., Ricinoleic acid in Vinca rosea seed oil, *J. Oil Technol. Assoc. India*, 19, 63, 1987.

13. Ahmad, M.U., Husain, S.K., and Osman, S.M., Ricinoleic acid in Phyllanthus niruri seed oil, *J. Am. Oil Chem. Soc.*, 58, 673, 1981.

14. Daulatabad, C.D., and Mirajkar, A.M., Ricinoleic acid in Artocarpus integrifolia seed oil, *J. Am. Oil Chem. Soc.*, 66, 1631, 1989.

15. Daulatabad, C.D., Desai, V.A., Hosamani, K.M., and Jamkhandi, A.M., Novel fatty acids in Azima tetracantha seed oil, *J. Am. Oil Chem. Soc.*, 68, 978, 1991.

16. Daulatabad, C.D. and Hosamani, K.M., Unusual fatty acids in Brunfelsia americana seed oil: a rich source of oil, *J. Am. Oil Chem. Soc.*, 68, 608, 1991.

17. Yunusova, S.G., Gusakova, S.D., Glushenkova, A.I., Lipids from the seeds of Securinega suffruticosa, *Khim. Prir. Soedin. (Tashk.)*, 3, 277, 1986.

18. Kleiman, R., Spencer, G.F., Earle, F.R., Nieschlag, H.J., and Barclay, A.S., Tetra-acid triglycerides containing a new hydroxy

eicosadienoyl moiety in Lesquerella auriculata seed oil, *Lipids*, 7, 660, 1972.

19. Sreenivasan, B., Kamath, N.R., and Kane, J.G., Studies on castor oil. 1. Fatty acid composition of castor oil, *J. Am. Oil Chem. Soc.*, 33, 61, 1956.

20. Badami, R.C., and Kudari, S.M., Analysis of Hiptage madablota seed oil, *J. Sci. Food Agric.*, 21, 248, 1970.

21. Nikolin, A., Nikolin, B., and Jankovic, M., Ipopurpuroside, a new glycoside from Ipomoea purpurea, *Phytochemistry*, 17, 451, 1978.

22. von Wettstein-Knowles, P.M., Waxes, cutin, and suberin, in *Lipid Metabolism in Plants*, T.S. Moore Jr., Ed., CRC Press, Boca Raton, 1993, 127-166.

23. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

24. Singh, I., and Jones, K.W., The use of heparin as a simple and cost-effective means of controlling background in nucleic acid hybridization procedures, *Nucleic Acids Res.*, 12, 5627-5630, 1984.

25. Frohman, M.A., Dush, M.K., and Martin, G.R., Rapid production of full-length cDNAs from rare transcripts: amplification using s single gene-specific oligonucleotide primer, *Proc. Natl. Acad. Sci. USA*, 85, 8998-9002, 1988.

26. Wada, H., Gombos, Z., and Murata, N., Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation, *Nature*, 347, 200-203, 1990.

27. Keegstra, K., Olsen, L.J., and Theg, S.M., Chloroplastic precursors and their transport across the envelope membranes, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 40, 471-501, 1989.

28. van de Loo, F.J., and Somerville, C., Plastid ω-3 fatty acid desaturase cDNA from *Ricinus communis*, *Plant Physiol.*, submitted.

29. Coppens, N., Biosynthesis of fatty acids by seeds of *Ricinus communis*, *Nature*, 177, 279, 1956.

30. Canvin, D.T., Formation of oil in the seed of Ricinus communis L., *Can. J. Biochem. Physiol.*, 41, 1879-1885, 1963.

31. Yamada, M., and Stumpf, P.K., Enzymic synthesis of ricinoleic acid by extracts of developing Ricinus communis L. seeds, *Biochem. Biophys. Res. Commun.*, 14, 165-171, 1964.

32. James, A.T. Hadaway, H.C., and Webb, J.P.W., The biosynthesis of ricinoleic acid, *Biochem. J.*, 95, 448-452, 1965.

33. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., Basic local alignment search tool, *J. Mol. Biol.*, 215, 403-410, 1990.

CHAPTER 5

LARGE SCALE SEQUENCING OF SEED-SPECIFIC CASTOR CLONES

ABSTRACT

Clones of a castor developing seed library were differentially screened for those not expressed in leaves, and not expressed in seeds at very high levels. Terminal (5') sequence data were obtained by automated DNA sequencing for 526 such clones. Of these ESTs generated, 468 contained informative sequence, and a putative identification could be made for 46% of them by comparison to sequence databases using the program Blastx.

Among the clones identified were two clones of a fatty acid desaturase homologue, which appeared to encode a novel desaturase or desaturase-like protein. These are considered putative oleate-12-hydroxylase clones.

In addition, a range of other novel plant genes were identified, and the possible functions of these genes in the developing castor seed are discussed.

INTRODUCTION

Recent advances in technology accruing from the increased interest in genome sequencing have made it possible to sequence DNA rapidly and voluminously using automated instruments. By automation, it has become possible to do experiments

involving large amounts of DNA sequencing, where previously these experiments were prohibitively demanding of investigator time. This technology was exploited in the continuing quest for a gene encoding the oleate-12-hydroxylase of castor.

The design of this large-scale DNA sequencing project was based upon three considerations. Firstly, ricinoleic acid is specific to the seed tissue of castor, and is not found in vegetative tissues or synthesized in germinating seeds.[1-4] A differential screening approach was therefore used to sequence only clones that were seed-specific. Differential screening has been used with success to isolate induced genes, particularly genes expressed in a given tissue following treatment with a drug, environmental variable, etc, but not expressed in the same tissue at the same developmental stage without treatment (for example, see Hajela *et al*[5]). In these cases a small number of induced sequences can be detected which may have direct relevance to the treatment. The differential screen used here is qualitatively different from the usual approach, in that the genes expressed in two completely different organs are compared, and a much larger proportion of sequences should differ between the two (discussed further below).

Secondly, it is hypothesized that mRNA for the oleate-12-hydroxylase is moderately abundant in the mRNA pool in developing endosperm. This is based upon the same considerations outlined above (chapter 3), namely that the enzyme is very active in this tissue, and probably does not have an extraordinarily high turnover number. In fact it is estimated that, to a first approximation (disregarding possible subtleties of gene regulation), the mRNA abundance should be similar to that for the

stearoyl-ACP-desaturase. These enzymes are involved in the same pathway, perform similar fatty acid oxidations and, it is hypothesized (chapter 4), that they do so with the same cofactor, hence might have similar turnover numbers. Clones for the stearoyl-ACP desaturase have been isolated from developing castor endosperm cDNA libraries at frequencies of 1/3000 (chapter 3) to 1/1250 (below). Furthermore, in this experiment identification of cDNA clones was made on the basis of their sequence, not by functional expression as in chapter 3. This considerably reduces the number of clones which must be examined in this experiment, since they need not be full-length or in a particular orientation. Indeed, partial cDNA clones were preferable, ensuring that DNA sequence obtained was from the coding sequence and hence more readily identified by comparison to databases, since coding sequences of genes are more highly conserved in evolution than non-coding sequences. The cDNA library to be used for the experiment was therefore first examined for the relative abundance of partial cDNA clones. In addition to moderately-abundant seed-specific clones, some seed-specific clones (viz. seed storage proteins) were expected to be highly abundant. These clones were eliminated from the experiment.

The third consideration was whether the oleate-12-hydroxylase DNA sequence would be recognisable as such. In chapter 4 it was hypothesized that this gene is homologous to those of the fatty acid desaturases, particularly *fad2*, the microsomal Δ12 desaturase gene. However, attempts to isolate the oleate-12-hydroxylase gene by probing with the *fad3* gene were unsuccessful, and others have found that *fad2* could also not be isolated by this approach. Furthermore, it was also not possible to isolate

the oleate-12-hydroxylase gene using an apparently conserved desaturase motif in a PCR approach. It is still hypothesized (for the reasons described in chapter 4) that the oleate-12-hydroxylase gene is homologous to those of the fatty acid desaturases, however some of the desaturases, including *fad3* and *fad7* which were available for these experiments, are too divergent from the oleate-12-hydroxylase to be used as probes for direct gene isolation or gene isolation by conserved-sequence PCR. Nevertheless, sequence similarity should be obvious if the sequences were to be compared.

Another aspect of this large-scale sequencing experiment was the likelihood of isolating other genes of interest to our laboratory or other laboratories, regardless of the outcome with respect to the oleate-12-hydroxylase gene. For example, the microsomal acyltransferases are active in this lipid-synthesizing tissue, including the seed-specific diacylglycerol acyltransferase, and it was of interest to isolate any of the genes encoding these (and many other) enzymes.

MATERIALS AND METHODS

Screening of λZAPST Library

Duplicate filters (Hybond N^+, Amersham) were lifted from plaques of the λZAPST library. DNA was fixed to the filters by placing them on filter paper moist with denaturing solution (0.5 M NaOH, 1.5 M NaCl; 5 min), neutralising solution

(0.5 M Tris-Cl pH 7.4, 1.5 M NaCl; 5 min), and 2 X SSC (0.3 M NaCl, 0.03 M Na-Citrate, pH 7.0). The filters were then air-dried, with no further fixation of the DNA. A ^{32}P-labelled probe was prepared from the insert of the stearoyl-ACP-desaturase clone pRcD1[6] by random priming[7] and purified of unincorporated nucleotides by ethanol precipitation in the presence of ammonium acetate.[7] The filters were prehybridised at 65°C for ~1 h in the hybridization solution (4 X SET (0.6 M NaCl, 0.12 M Tris-HCl pH 7.4, 8 mM EDTA), 0.1% sodium pyrophosphate, 0.2% SDS, 5% dextran sulphate, and 0.1% heparin) before addition of the probe and hybridization overnight at 65°C. The filters were washed three times in 2 X SSC, 0.1% SDS at room temperature, then exposed to X-ray film.

Differential Screening

Phage in nine separate 96-well plates were replicated onto a single bacterial lawn the size of one 96-well plate using a 96-prong device which could be lowered onto the lawn through a 3 x 3 array of guides. The blunt ~1 mm diameter prongs carried sufficient phage to give plaques of consistent size, without significant encroachment between neighbouring plaques. Triplicate filters were lifted from these plaques and screened as described above, except that for plates 1-9, polyadenylic acid (1 μg ml^{-1}) was added to the hybridization solution, and results for plates 1-9 were obtained from a phosphor-imager (Molecular Dynamics) rather than from autoradiographs. Exposure times were: plates 1-9, 21 h (note that phosphor imaging

is several-fold more sensitive than autoradiography); plates 28-36 leaf probe 3 days, seed probe 24 h, redundant-clone probe 1.5 h.

Probes for screening plates 1-9 were prepared as follows.

Poly(A)$^+$ RNA (1 μg) from seed or leaf (chapter 3) in a volume of 17 μl was heated to 70°C for 5 min, then chilled in ice-water, and added to the reaction tube, to a final volume of 50 μl:

RNasin (Promega), 1 U μl^{-1}

reverse transcriptase buffer (Boehringer-Mannheim), 1 x

oligo$(dT)_{12\text{-}18}$, 20 ng μl^{-1}

dGTP, dATP, dTTP, 1 mM each

dCTP (unlabelled), 4.8 μM

α-^{32}P dCTP (3000 Ci mmol^{-1}), 100 μCi

Avian Myeloma Virus Reverse Transcriptase, 40 U.

The reaction was incubated at 42°C for 60 min. The reaction was stopped and RNA removed by addition of EDTA (to 16 mM), SDS (to 0.4%), NaOH (to 0.4 M) and incubation at 65°C for 30 min. The probe was neutralised with 6 μl 2 M HCl and 20 μl 1 M Tris-Cl, pH7.4, then precipitated with 375 μl EtOH in the presence of 0.7 M ammonium acetate and 10 μg denatured carrier (salmon sperm) DNA. After incubation at -20°C for ~3 h, DNA (~60% of total radioactivity) was pelleted by centrifugation for 15 min, and resuspended in 200 μl water and added to the filters.

For plates 28-36, first-strand cDNA was made using the same RNA (0.5 μg seed, 1.2 μg leaf) in a reverse transcription reaction similar to that described above,

but using unlabelled nucleotides and all other components from a reverse transcription kit (Promega). The RNA was hydrolysed and the cDNA was neutralised as described above, and then purified by batch chromatography on glass (GeneClean, Bio101). The cDNA was then used as a template for random priming,[7] using 100 μCi α-^{32}P dCTP. The probes were precipitated as described above, and heated to 100°C (5 min) before addition to the filters. Incorporation of radioactivity was ~60% (leaf probe) or ~30% (seed probe).

A probe was made from redundant clones as follows. Clones (Table 7) were digested with *Bam*HI and *Kpn*I and the inserts purified from agarose gels. DNA of these inserts was pooled and ~600 ng labelled with 100 μCi α-^{32}P dCTP (~80% incorporation) by random priming as described above.

DNA Sequencing

Plasmid DNA was prepared from *Escherichia coli* cultures (5 ml, LB medium containing 100 mg l^{-1} ampicillin) using "Magic Minipreps" (Promega) according to the instructions of the manufacturer, and submitted for automated cycle sequencing on Applied Biosystems 373A instruments using the fluorescent primer T3 (occasionally T7).

RESULTS AND DISCUSSION

Characterization of λZAPST cDNA Library

Bacterial colony hybridizations frequently give high background signal due to the bacterial debris adhering to the filter. Since a differential screen was planned, it was imperative to minimise background signal, and so a lambda phage library was chosen. The library used in this experiment (λZAPST) was constructed in λZAP II (Stratagene) by Dr. S. Turner, from developing endosperm and embryo of castor, as described above for the pYES2.0 library (chapter 3). Construction of the λZAPST library included directional cloning, so that 5' ends of the inserts should be found at the T3 side of the polylinker. This library had not yet been characterised, so an investigation was made similar to that described in chapter 3. The castor stearoyl-ACP-desaturase[6] probe was used to isolate 16 primary duplicated positive clones from a single filter bearing ~20 000 plaques. All 16 of these clones re-screened, indicating an abundance of about 1 per 1250, which is the same order of magnitude as for the pYES2.0 library (1 per 3000). DNA of 14 of the clones was analysed by restriction digestion and agarose gel electrophoresis. Full-length clones should contain an insert of ~1.6 kb giving a *Bam*H1 fragment of ~0.7 kb. Three of the 14 clones analysed appeared to be close to full-length. Inserts of the remaining clones varied between 0.5 kb and 1.0 kb.

Table 7. Redundant sequences obtained in batch 1 sequences, used to make a probe for screening of batch 2. Frequency of these sequences in batch 2 (after the screen) is compared with batch 1 (before the screen) to indicate effectiveness of the screen.

Redundant Sequence	Frequency in Batch 1	Frequency in Batch 2
Ribosomal proteins[1]	12 (6.6%)	23 (8.0%)
12S seed storage protein[2]	10 (5.5%)	8 (2.8%)
2S seed storage protein[3]	6 (3.3%)	5 (1.7%)
Heat shock proteins[4]	4 (2.2%)	5 (1.7%)
Enolase[5]	4 (2.2%)	0

1 Clones pCRS262, pCRS312, pCRS356, pCRS358, pCRS377, pCRS396, pCRS407, pCRS409, pCRS416, pCRS426, pCRS432, pCRS442, pCRS446.

2 Clones pCRS267, pCRS269, pCRS298, pCRS404, pCRS405, pCRS408, pCRS434, pCRS443, pCRS453, pCRS454.

3 Clones pCRS281, pCRS328, pCRS337, pCRS362, pCRS375, pCRS431.

4 Clones pCRS264, pCRS348, pCRS397.

5 Clones pCRS330, pCRS380, pCRS415, pCRS439.

Differential Screening with Seed and Leaf First-Strand cDNA Probes

Plaques of the λZAPST library were picked from fresh, low-density plates (to avoid cross-contamination) into 96-well plates. They were then plated onto bacterial lawns with the addition of several control clones. These included non-recombinant phage (which gave blue plaques on plates containing IPTG and XGal indicating that the *lacZ* gene was uninterrupted) and clones of the stearoyl-ACP-desaturase, acyl carrier protein, 2S seed storage protein, and 12S seed storage protein, the latter three clones being provided by Dr. S. Turner. By using an offsetting device, plaques from nine 96-well plates could be represented on a filter the size of one plate. Multiple filters, each representing 864 identifiable clones, were lifted from the resulting plaques and screened with ^{32}P-labelled first-strand cDNA probes reverse transcribed from leaf or developing endosperm/embryo poly(A)$^+$ RNA. Only those clones were selected which gave no detectable signal with the leaf probe, and did not give a very strong signal with the seed probe. These results were obtained by on-screen analysis of images from a phosphor-imaging system (Molecular Dynamics). Plates 1-9 were processed in this manner, from which the first batch of sequences were obtained.

Of 864 possible plaques from plates 1-9, 10 did not appear and 15 were observed to be occluded by bubbles separating the plaque and filter, leaving 839 clones with DNA on the filter. Of these, 162 (19.3%) were scored as having a strong seed signal, while 280 (33.4%) gave no detectable signal with the leaf probe. Of these 280, 222 were not among the previous category and were selected for

sequencing. These results therefore indicated that 222 of 839, or 26.5% of clones, were in the category "seed-specific and not highly abundant". Of the 162 clones having a strong seed signal, only 58 appeared to be seed specific.

Some changes were made when screening plates 10-54 for the second sequencing batch. The seed and leaf probes were made by random priming using first-strand cDNA as a template, in an attempt to gain maximum incorporation of radioactivity into less-abundant sequences. In addition, a probe was made from the pooled insert DNA of clones frequently sequenced in the first batch (Table 7) so that fewer redundant sequences would be obtained. Screening results were obtained directly from autoradiograms. The most clearly-interpretable data were for plates 28-36, from which clones for the second sequencing batch were isolated. These autoradiograms are presented in Figures 12-14.

For plates 28-36, 851 of a possible 864 plaques were represented on the filter, and of these 851, 370 (43.5%) gave a strong seed signal, 512 (60.2%) gave no detectable leaf signal, and 141 (16.6%) gave a signal with the probe made from redundant sequences (the effectiveness of screening with this particular probe is discussed below). This resulted in the selection of 348 (40.9% of 851) clones to be sequenced.

Considerable differences are evident among the results of the first and second screening batches. More (43.5% compared with 19.3%) clones were classed as having a strong seed signal in the second batch than in the first. This may be attributable to the subjective nature of determining when the signal was strong as

Figure 12. Autoradiogram of plates 28-36 probed with ^{32}P-labelled cDNA from developing seeds. The positions of the wells in the original 96-well plates is indicated (borders), and the position of clones from each 96-well plate relative to other 96-well plates is indicated in the box.

Figure 13. Autoradiogram of plates 28-36 probed with ^{32}P-labelled cDNA from developing leaves. The positions of the wells in the original 96-well plates is indicated (borders), and the position of clones from each 96-well plate relative to other 96-well plates is indicated in the box.

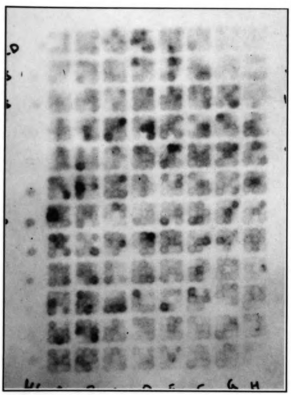


Figure 14. Autoradiogram of plates 28-36 probed with ^{32}P-labelled DNA from redundant clones sequenced in batch 1. The positions of the wells in the original 96-well plates is indicated (borders), and the position of clones from each 96-well plate relative to other 96-well plates is indicated in the box.

opposed to merely detectable, and the inevitable differences in background obtained with different probes and filters in separate hybridisations. In particular, probes for the second batch were made by random priming off first strand cDNA on the premise that DNA polymerase I (Klenow) is less sensitive to low concentrations of the labelled nucleotide than is reverse transcriptase, and would therefore give greater incorporation of label. This may have been a successful strategy for the seed probe, but apparently not for the leaf probe, with which only 39.8% of clones gave a detectable signal in the second batch, compared with 66.6% in the first batch.

Despite the differences between the two batches, these results are in general agreement with those of RNA excess/single-copy DNA hybridization experiments in tobacco.[8,9] These experiments show that the total complexity of mRNA or nuclear RNA in different organs of the tobacco plant is very similar. The organs analysed in these experiments were leaf, root, stem, ovary and anther, but the similarity of the results for these organs suggest that it is possible to extrapolate to other organs also, such as the seed. Approximately 25000 mRNAs were expressed in each organ. Of these, approximately 25-30% were ubiquitous, that is, expressed in all organs, while an additional 30-40% were expressed in another organ (but not all other organs). The remaining 30-45% of messages expressed in a particular organ were unique to that organ.[8]

Sequencing and Sequence Analysis

The differential screens described above gave a total of 570 lambda phage clones selected for sequencing. These were individually excised *in vivo* to yield the corresponding Bluescript plasmid. DNA was prepared from each plasmid, analysed spectrophotometrically for DNA concentration, and submitted to the PRL Plant Biochemistry Facility for automated sequencing. The T3 primer was used because in the directional library, this should give sequence from the 5' end of the clones, and since most of them are not full-length, this should frequently yield protein-coding sequence which is most readily identified by database searches. In some sequences (e.g. CRS262, CRS263, CRS269), a poly(A) tail was obvious, indicating that the clone was small enough to sequence through in one run. In fewer cases (e.g CRS270, CRS294, CRS315), a poly(T) tract was observed at the 5' end of the insert DNA, indicating that in some cases directional ligation fails during library construction. For this reason, sequences which apparently gave database matches in a negative reading frame were not ignored. Sequence data was edited to remove vector/linker sequences, and truncated at the point where sequence quality declined substantially. These edited sequences (typically 400-500 nucleotides) were submitted electronically to the National Center for Biotechnology Information, Bethesda, MD. DNA sequences were compared in all reading frames to the non-redundant translated-nucleotide and protein sequence databases (Swiss-Prot 24.0 or 25.0 plus weekly updates; PIR 35.0, 36.0, or 37.0; GenBank Release 75.0, 76.0, or 77.0, plus daily

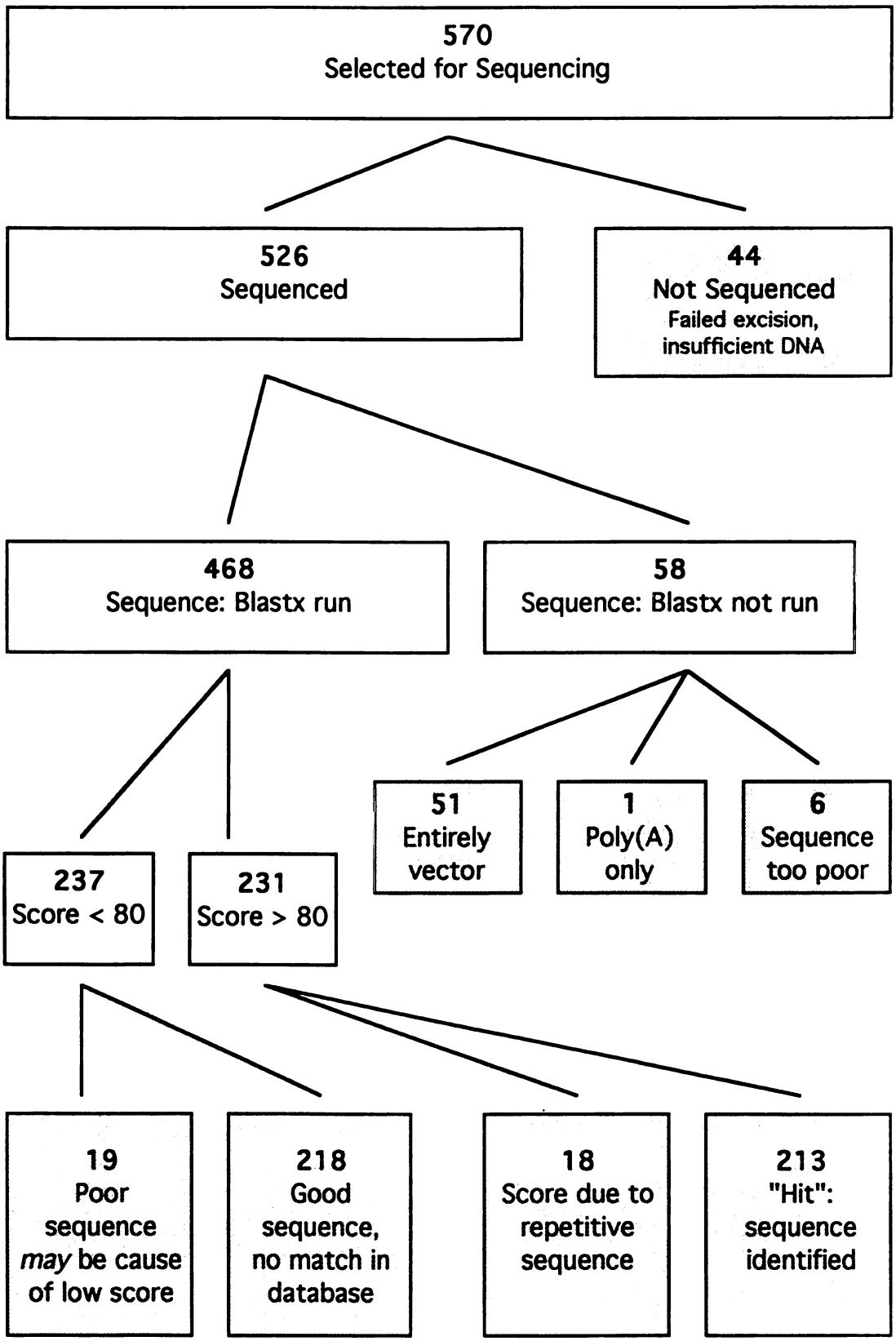
updates; and EMBL Release 34.0 or 35.0, plus daily updates) by the program blastx,[10] in the months March-July, 1993. In general, a blastx score of 80 is used as the cutoff for determining whether a sequence has a match in the databases. An exception is made for repetitive sequences, which may give scores higher than 80 without meaningful identification.

In addition to the 570 clones selected for sequencing on the basis of the differential screen, five clones were sequenced which gave a strong signal with the seed probe and would therefore normally have been excluded. Four encoded 2S seed storage protein, and the fifth encoded 12S seed storage protein. This confirmed that the clones excluded on the basis of strong seed signal were not of interest in identifying novel genes.

Data in Table 7 indicate that screening the second batch to remove redundant clones obtained in the first batch was partially successful. The frequency of seed storage protein clones sequenced in the second batch is lower, but not zero, probably due in part to the weak signal given by clones with very short inserts. No enolase clones were sequenced in the second batch. There may have been a small reduction in the frequency of heat shock protein clones, but there was no effect on the frequency of clones for ribosomal proteins. This is not surprising since these were generally not truly redundant clones, most clones corresponding to different heat shock proteins or ribosomal proteins.

Figure 15 is a flow diagram depicting the various fates of the total of 570 clones selected for sequencing. Insufficient DNA of 44 clones was obtained (in a

Figure 15. Fates of clones selected for sequencing by differential screening.


570
Selected for Sequencing
526
Sequenced
44
Not Sequenced
Failed excision,
insufficient DNA
468
Sequence: Blastx run
58
Sequence: Blastx not run
51
Entirely
vector
1
Poly(A)
only
6
Sequence
too poor
237
Score < 80
231
Score > 80
19
Poor
sequence
may be cause
of low score
218
Good
sequence,
no match in
database
18
Score due to
repetitive
sequence
213
"Hit":
sequence
identified

single attempt) for sequencing. Of the remaining 526 clones sequenced, 58 gave sequence data which was not considered informative, containing, for example, extraneous DNA. Sequence from the 468 informative clones was analysed by the blastx program, leading to the putative identification of 213 (46%) of them. This is a high level of identification, compared to other similar sequencing projects. Of 2375 human brain cDNAs, 17% were identifiable by comparison to the databases.[11] In an earlier report of human brain cDNA sequences where less stringent criteria were used for the assignment of identity, 52% of 475 cDNAs were identified.[12] Of 130 random maize leaf cDNAs sequenced, 20% were identified.[13] The identification of only 8% of 830 cDNAs from rice suspension culture cells was reported,[14] but a much higher alignment score (160, compared with 80 used in this and other projects) was used as the criterion for identification. Differential screening approaches have also been used previously in this type of sequencing project. In one study,[15] mouse testis cDNAs were screened with probes made from testis and from a pool of other tissues. The majority (90%) of 51 abundant clones which hybridised with both probes were identified, but only 25% of 120 rare clones, which hybridised with neither probe, were identified.

Sequences for which an alignment with database sequences suggests an identity (blastx score $\geq$ 80) are presented in four tables. In Table 8 are listed those clones for which the most similar match in the databases was a prokaryotic gene. Similarly, Table 9 lists clones for which the most similar match was from a eukaryote other than a higher plant. Clones for ribosomal protein genes have been excluded from this

table and are listed in Table 10. In Tables 8 and 9, the possible function of each clone is discussed, in an attempt to reveal new opportunities arising from the identification of genes apparently not previously sequenced from higher plants.

The fourth table, Table 11, lists most of the clones in the remaining category, by the other higher plant gene to which they were found to be most similar. Clones for ribosomal proteins are listed separately, in Table 10. Clones in Table 11 are grouped according to the general area of metabolism or cell function with which they are most likely to be involved, and these categories are discussed below, rationalising the sequences obtained in the light of the major metabolic or cellular processes known to be active in the developing seed tissue from which the cDNA library was made, and for which the differential screen employed might be expected to enrich clones.

Table 8. Clones for which the most similar match in the databases was from a prokaryote. The clones are listed by the suffix of their clone number (pCRS265 etc.). Information for each clone is structured as follows: blastx score, (probability of random alignment), database (G: GenBank; S: SwissProt; P: PIR), accession number, protein description, organism. This is followed by a discussion of the possible function of each clone. Information for these discussions is drawn from annotations to the database entries, as well as standard textbooks, and original references where cited.

---

265 Blast 82 (0.00019) S P04990: threonine synthase, *Bacillus subtilis*.
18/39 (46%) residues identical between CRS265 and *B. subtilis* protein, 28/39 (71%) when allowing for conservative substitutions. Residues conserved between the *E. coli* and *B. subtilis* enzymes[16] are generally also conserved in the castor sequence, suggesting that the marginal score reflects a true identification. Enzymes of amino acid biosynthesis must be active in the castor endosperm for storage protein synthesis.

296 Blast 418 (1.3e-55) S P17242: asparaginyl-tRNA synthetase, *E. coli*.
Also a weaker match to the yeast homologue (43/78 = 55% identical compared with 81/125 = 64% identical in *E. coli*). Tyr-426 of the *E. coli* protein is involved in ATP binding[17] and is conserved in the castor sequence. This is in a region which is strongly similar between the two proteins, and is also conserved between other aminoacyl-tRNA synthetases for amino acids with an XAX codon (includes Tyr, stop, His, Gln, Asn, Lys, Asp, and Glu).[18] Castor 2S seed storage protein is notably rich in glutamine (44 residues of 258), raising the possibility that this clone is in fact a glutaminyl-tRNA synthetase.

332 Blast 119 (1.4e-09) S P26242: hypothetical 67.9 kDa protein in pufx region (photosynthetic gene cluster), *Rhodobacter capsulatus*.
No other matches. *Rhodobacter* sequence has similarity with enzymes requiring thiamine pyrophosphate, such as acetolactate synthase, pyruvate decarboxylase and indolepyruvate decarboxylase, and transketolases. the predicted *Rhodobacter* protein has a slight mean hydrophilicity, suggesting that it is not a membrane protein. Only mutations affecting the differentiation of the photosynthetic aparatus are known to map to this gene cluster in *Rhodobacter*, suggesting that the predicted protein has some role in bioenergetics.[19]

336 Blast 153 (9.0e-14) S P22256: 4-aminobutyrate aminotransferase (GABA transaminase), *E. coli*.
Other matches are also bacterial amino acid metabolism enzymes (2,2-

dialkylglycine decarboxylase, 3-isopropylmalate dehydrogenase, acetylornithine aminotransferase). In each case, sequence similarity is found in the same 2 or 3 small clusters, presumably of functional significance. The same applies also for two weaker matches in the same reading frame to the yeast enzymes ornithine aminotransferase and 4-aminobutyrate aminotransferase. Comparison of the GABA transaminase with several other aminotransferases also revealed these conserved domains, but their functional significance has not been elucidated.[20] Note: CRS841 has strong homology to alanine aminotransferase of millet, human and rat, as well as some similarity to aspartate aminotransferase of *Rhizobium*. CRS856 has strong homology to aspartate aminotransferase from higher plants and other organisms.

435 Blast 106 (1.4e-07) G D10483: ORF, *E. coli*.
Also weaker matches in the same reading frame to rat and human short-chain specific acyl-CoA dehydrogenases, and the residues which match between these and castor are also those most strongly conserved amongst various rat and human acyl-CoA dehydrogenase isozymes,[21] and are generally the same as residues matching with the *E. coli* sequence. Hence probably a true match for an acyl-CoA dehydrogenase. The rat and human proteins are mitochondrial flavoprotein enzymes for fatty acid β-oxidation (butanoyl-CoA → 2-butenoyl-CoA). The developing castor seed may be "preparing" for germination by making glyoxisomes (cf. CRS266, Table 9) and the enzymes of ß-oxidation they require.

463 Blast 117 (4.3e-09) S P08534: NIFM protein, *Klebsiella pneumoniae*.
Also matches to NIFM of *Azotobacter* spp., and to protein export protein PRSA of *Bacillus subtilis*, in the same reading frame and at the same clusters of conserved residues, suggesting that NIFM and PRSA and CRS463 all have a related function. NIFM is probably involved in a processing step necessary to generate an active nitrogenase iron protein. PRSA is involved in a late stage of protein export and is probably located on the outer membrane. An hypothetical role of CRS463 might involve storage protein import into the vacuole.

519 Blast 118 (4.4e-21) S P00968: carbamoyl phosphate synthase large (ammonia) chain, *E. coli*.
Also matches to other bacterial, mammalian and yeast carbamoyl-phosphate synthases, some the ammonia-dependent enzyme involved in arginine biosynthesis, some the glutamine-dependent enzyme involved in pyrimidine biosynthesis. In liver mitochondria, carbamoyl phosphate synthase I (ammonia dependent) is tightly associated with ornithine carbamoyl transferase (CRS772), forming a multifunctional protein which can most efficiently use the labile carbamoyl phosphate for synthesis of citrulline (precursor to arginine). Presumably the arginine biosynthetic pathway is active in the developing castor

seed for storage protein synthesis. Alternatively, arginine synthesis may be active in the developing castor seed to support synthesis of polyamines (discussed in the text).

559 Blast 134 (5.0e-12) S P23523: hypothetical 31.0kDa protein in rnpB 3' region (ORF2), *E. coli*.
No other matches. The function of this *E. coli* ORF is unknown, but it shows weak (approx 24%) amino acid sequence similarity with *gntZ* of *Bacillus subtilis*, which affects the utilization of gluconate, and *gnd* of *E. coli*, which encodes 6-phosphogluconate dehydrogenase.[22] No plant gene for this enzyme was found by searching the databases. 6-Phosphogluconate dehydrogenase is an enzyme of the pentose phosphate pathway, which may[23] or may not be[24] important in providing NADPH for fatty acid synthesis.

631 Blast 162 (4.1e-17) G X51510: riboflavin synthase, *B. subtilis*.
Matches also to other bacterial riboflavin synthase β subunits. Riboflavin synthase is itself a flavoprotein, involved in the last steps of riboflavin synthesis. (Mammals cannot make riboflavin).

640 Blast 100 (1.6e-06) P S18956: fix23-4 protein, *Rhizobium meliloti*.
Fix23 is a multi-ORF locus bearing homology to fatty acid synthases and polyketide synthases, involved in synthesis of a secreted lipopolysaccharide.[25] Other matches are to mycocerosic acid synthase, a multifunctional protein of *Mycobacterium tuberculosis* which catalyses fatty acid elongation,[26] the polyketide synthase multifunctional protein of *Saccharopolyspora erythraea* involved in erythromycin biosynthesis,[27] and the type-I fatty acid syunthases of chicken and rat. Fix23-4 has homology to enoyl reductase, and CRS640 matches the conserved residues in the NAD(P)H binding site. There is also homology to alcohol dehydrogenases and zeta-crystallins (quinone oxidoreductase homologues) in the same region. CRS640 appears to align to the enoyl-reductase domain of the other matches also, but has no homology to the enoyl-ACP reductase cloned from rapeseed that is a component of the plant plastid type-II fatty acid synthase. No genes for plant type-I fatty acid synthases are yet cloned. Potentially CRS640 is from a cytoplasmic fatty acid elongase or polyketide synthase. Further sequence has been obtained: from the 3' end, and by sequencing further from the 5' end, thus covering the entire insert. The assembled sequence now has a blast score for fix23-4 of 113 (8.0e-08). The lack of sequence matching a different domain of a FAS suggests that this cDNA is from a gene encoding only one domain/protein of a putative FAS (though other genes may be found in the same region of the chromosome).

673 Blast 226 (1.4e-27) S P21773: hypothetical 17.4kDa protein in firA (ssc)-lpxA intergenic region, *Salmonella typhimurium*.

Match also to the homologous ORF in *E. coli*, and a shorter alignment (hence much weaker score: 61, 13/29 identical residues) in the same reading frame to 3-hydroxydecanoyl-ACP-dehydratase also of *E. coli*. However, the residues identical between CRS673 and the 3-hydroxydecanoyl-ACP-dehydratase are notably the same as those conserved between CRS673 and the *Salmonella* and *E. coli* ORFs in the same region, suggesting that these ORFs and CRS673 have something functional in common with the FAS enzyme.

The E. coli 17.4 kDa ORF is found in an operon which includes other ORFs and the two known genes of lipid A synthesis, *lpxA* and *lpxB*.[28-30] The 17.4kDa ORF is immediately upstream of *lpxA*, with overlapping start and stop codons, suggesting cotranscription and translational coupling.[28] Lipid A is an essential component of the outer leaflet of the outer membrane of most gram-negative bacteria.[31] It has a glucosamine backbone rather than the glycerol backbone of typical membrane lipids, and acyl chains are attached by ester and amide linkages.[31,32] The primary acyl chains are 3-hydroxy-myristic acid, to the hydroxyl groups of which are esterified the secondary acyl chains, lauric or myristic acids. In the *firA* (*omsA*) mutant of *E. coli*, the secondary acyl groups include palmitic acid, but the function of the *firA* gene product is unknown.[33] The first enzyme in lipid A synthesis, UDP-*N*-Acetylglucosamine 3-*O*-Acyltransferase, is encoded by *lpxA*.[28,34] The gene *lpxB* encodes lipid A disaccharide synthase.[29,35] The genes encoding other enzymes of lipid A synthesis have yet to be identified,[31,33] though the ORFs associated with *lpxA* and *lpxB* are obvious candidates. Other enzymes required for lipid A synthesis[31] include a deacetylase (hydrolysing acetate from the glucosamine group); a 3-hydroxymyristoyl-ACP *N*-acyltransferase; and the "late" acyltransferases, specific for lauroyl- and myristoyl-ACPs, which esterify these acyl groups to the hydroxyl groups of the previously esterified 3-hydroxy-myristate moieties. These "late" acyltransferases have been assayed in crude *E. coli* fractions, but have not been purified or well characterized.[36,37]

The FAS enzyme, 3-hydroxyacyl-ACP-dehydrase, is involved (possibly together with an isomerase) in the introduction of a double bond in the growing acyl chain, enabling the production of unsaturated fatty acids under anaerobic conditions. A single histidine (His^{70}) is the only active site residue that is directly involved in the reaction of this enzyme,[38] and this residue is also found in CRS673 and the two *E. coli* and *Salmonella* 17.4 kDa hypothetical proteins. 3-hydroxyacyl-ACP-dehydrase has been purified from spinach and has two histidine residues per molecule[39], as does the *E. coli* enzyme, but has not been cloned from higher plants.

The possible homology between the 3-hydroxyacyl-ACP-dehydrase and the 17.4 kDa ORFs suggests a possible function for the ORFs. The "late" acyltransferases produce estolide-linkages, where one acyl group is esterified to the hydroxyl-group of a substituted acyl chain. In *Claviceps*, such estolide linkages are formed by addition of the secondary acyl group to a double bond of the primary acyl group (see chapter 2). It is possible that a component of

the "late" acyltransferases is a 3-hydroxyacyl-dehydrase, giving the 2,3-unsaturated primary acyl group, to which the secondary acyl group is then added.

A report[40], of which I was only able to obtain the abstract at time of writing, suggests that the 17 kDa ORF described here was purified as a component of an "actomyosin" complex of *E. coli*.

706 Blast 266 (1.7e-32) G L14862: biotin carboxylase, *Anabaena* sp. (PCC 7120).
Biotin carboxylase is one of the two subunits of the *Anabaena* acetyl-CoA-carboylase (ACCase), carboxylating the carrier protein, before the carboxyl is transferred to acetyl-CoA. There are also matches to the biotin carboxylase subunits of other bacterial ACCases, and weak matches to the biotin carboxylase domain of yeast pyruvate carboxylase and various mammalian carboxylases. The *Anabaena* match (50/73 = 68% identical) is considerably stronger than the next strongest, *E. coli* (34/72 = 47% identical). This makes it tempting to suggest that this is a plastid enzyme. Dr. Ohlrogge compared CRS706 to their plant ACCase sequences, and found that this clone is significantly different, and so would be some other biotin carboxylase. However, it is possible that the previously cloned plant eukaryotic-type ACCase genes are for a cytoplasmic isozyme, since no transit peptide was identified. Such a cytosolic enzyme may produce malonyl-CoA for the synthesis of flavonoids, phytoalexins, ethylene, or fatty acid elongation (see CRS640).[23] Dr Ohlrogge is interested in the possibility that pCRS706 encodes a subunit of the plastid form of ACCase, responsible for provision of malonyl CoA for fatty acid synthesis, and will characterise this clone in more detail. There is evidence for a pokaryotic-type ACCase in plastids, and one subunit (of the carboxyltransferase) has been identified in pea chloroplasts.[41]

756 Blast 144 (7.0e-13) S P05054: ribokinase, *E. coli*.
Weaker matches to tagatose-6-phosphate kinase of *Staphylococcus aureus*, 1-phosphofructokinase of *Rhodobacter capsulatus* and *E. coli* and *Xanthomonas campestris*, a probable ribokinase of yeast, and ketohexokinase of rat. Clearly CRS756 encodes some carbohydrate kinase. Note also CRS517: very strong homology to 6-phosphofructokinase (β-subunit) from potato; CRS792: fructokinase, potato; CRS304: pyruvate kinase.

772 Blast 189 (1.4e-21) S Q02095: ornithine carbamoyltransferase, anabolic, *Mycobacterium bovis*.
Similar matches to a number of other bacterial ornithine transferases, and weaker matches to the enzyme from various mammals. Arginine biosynthesis. See CRS519.

798 Blast 178 (2.7e-17) P A42653: pyruvate dehydrogenase E1-α subunit,

Acholeplasma laidlawii.
Matches also to homologues from bacteria and mammals (*Acholeplasma* is a mycoplasma). Part of PDH complex. There are E1 α and β subunits, these are distinguished from the dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase components of the complex. CRS504 and CRS844 are homologous to dihydrolipoamide dehydrogenase (see text). In the developing castor seed, the source of carbon for fatty acid biosynthesis has been proposed to be from imported sucrose via glycolysis to malate in the cytoplasm, followed by the action of malic enzyme and pyruvate dehydrogenase in the plastid, yielding acetyl-CoA and NAD(P)H.[23] Components of pyruvate dehydrogenase are therefore logically abundant sequences expressed in the developing castor seed.

828 Blast 189 (5.0e-20) S P31038: succinate dehydrogenase flavoprotein subunit, *Rickettsia prowazekii.*
Also matches to the same protein from yeast, other bacteria, and mammals, as well as weaker similarity to bacterial fumarate reductase flavoprotein. In bacteria, succinate dehydrogenase is a complex containing the flavoprotein, an iron-sulphur protein, and cytochrome b_{558}. In mammals, the enzyme has two subunits, the larger being the flavoprotein. Succinate dehydrogenase oxidises succinate to fumarate in the tricarboxylic acid cycle.

852 Blast 145 (6.4e-14) S P80016: nitrogen-regulatory protein P-II, *Synechococcus* sp. (PCC6301).
Also matches to the homologous protein from a range of bacteria. Involved in deadenylation (activation) of glutamine synthase under nitrogen-limiting conditions (when ratio of glutamine to 2-ketoglutarate decreases). P-II is also involved in the regulation of the transcription of the glutamine synthase gene. P-II is controlled by photosystem II.[42] Regulation of glutamine synthase in plants by the same mechanisms as in these bacteria is hitherto unknown.[43] CRS852 may be involved in regulation of amino acid/nitrogen metabolism in the castor seed, where amino acid and protein synthesis has to be coordinated with nitrogen import. P-II is a small (112 amino acid) protein, and the alignment with CRS852 extends from the N-terminus, suggesting that pCRS852 may be a full length clone. A collaboration has been initiated with Dr. Gloria Coruzzi, New York University.

Table 9. Clones for which the most similar match in the databases was from a eukaryote other than a higher plant. Ribosomal proteins are not included. The clones are listed by the suffix of their clone number (pCRS266 etc.). Information for each clone is structured as follows: blastx score, (probability of random alignment), database (G: GenBank; S: SwissProt; P: PIR), accession number, protein description, organism. This is followed by a discussion of the possible function of each clone. Information for these discussions is drawn from annotations to the database entries, as well as standard textbooks, and original references where cited.

---

266 Blast 151 (5.6e-15) S P14292: 20kDa peroxisomal membrane protein, *Candida boidinii*.
Membrane-associated protein of methanol-induced peroxisomes. "Function very likely to be related to the metabolism of methanol"[44] presumably would not explain its occurrence in castor seed. Targetting of proteins to peroxisomes appears to be broadly conserved throughout the eukaryotes, and the C-terminal tripeptide (S/A/C)-(K/R/H)-L may be sufficient to direct peroxisomal protein import in plants, yeast, insects and mammals.[45] This consensus sequence is common among peroxisomal proteins, including plant peroxisomal soluble proteins,[46] but is not the only type of peroxisomal targetting determinant.[47] The *Candida* protein has the C-terminal tripeptide A-K-L. The sequence CRS266 aligns to the central portion of the *Candida* sequence; a second sequencing run, from the 3' end of the clone, revealed that the open reading frame terminates with the variant tripeptide K-A-L. Since this is the first plant putatively peroxisomal membrane protein sequenced, it is not clear if this variant tripeptide could function in targetting to the plant peroxisome. A comparison of C-terminal sequences of a number of yeast peroxisomal proteins revealed that they are rich in hydroxy amino acids at the -8 and -9 positions;[44] serine occupies both these positions in the castor sequence. The developing castor seed may make glyoxisomes in "preparation" for germination when glyoxisomal function dominates metabolism (see also CRS435). A collaboration has been initiated with Dr. Laura Olsen (University of Michigan) who is interested in using the clone as a possible marker for plant peroxisome biogenesis.

282 Blast 297 (8.2e-37) S P20659: trithorax, *Drosophila melanogaster*.
A 404 kDa nuclear zinc-finger protein involved in segment determination by interacting with the genes of *bithorax* and *antennapedia*.[48] Homology is to the carboxy terminus, not surprising for such a large protein/message. Trithorax mutants have homoeotic segmentation. Also a match for the human trithorax homologue, the function of which is not known, but is implicated in translocations resulting in leukemias,[49] i.e. seems to have a DNA-binding role. Possibility of controlling embryo-development in castor? CRS282 would be

interesting to map in *Arabidopsis* where several interesting mutants affecting embryo development are mapped, and to try to correlate with a function by antisense-inhibition in *Arabidopsis* upon isolating the *Arabidopsis* homologue.

304 Blast 165 (6.3e-16) S P00548: pyruvate kinase, chicken muscle.
Final step in glycolysis (phospho-enol-pyruvate → enol-pyruvate →spontaneously pyruvate), hence high level of expression in castor seed would not be surprising, to maintain the acetate pool used for fatty acid synthesis. However, it has been suggested that the major source of pyruvate in the plastids of developing castor endosperm is not made via pyruvate kinase, but rather from oxaloacetate which is converted to malic acid by malate dehydrogenase (CRS503) and to pyruvate by malic enzyme.[23] Nevertheless, when measured *in vitro*, the leukoplasts of the developing castor endosperm have high activities of pyruvate kinase (330 nmol/min/mg protein) and some phosphoglycerate kinase (9 nmol/min/mg protein). These enzymes are proposed to participate in a triose-phosphate shuttle (also requiring malate dehydrogenase, CRS503, and glyceraldehyde-3-phosphate dehydrogenase, CRS294) which contributes to the provision of ATP for fatty acid biosynthesis in non-photosynthetic tissues such as the castor seed.[24] This shuttle would be particularly important if ATP is not derived from glycolysis in the plastid, as in the case of the malic enzyme proposal, or imported from the cytoplasm (see CRS692). A cytosolic pyruvate kinase has been cloned from potato, but this was the weakest match against CRS304 compared to all the other non-plant pyruvate kinases, so CRS304 is quite divergent from the potato clone. Dr. David Dennis, Queens University (personal communication 4/8/93) has isolated at least four pyruvate kinase genes from plants (not available in the database), and believes there may be more, quite divergent ones. This clone might be an example of such a gene.

314 Blast 84 (7.3e-05) P S21342: ring-infected erythrocyte surface antigen, *Plasmodium falciparum*.
Alignment: 18/43 (41%) identities, 28/43 (65%) positives.

347 Blast 189 (4.9e-22) G S51858: MO25 gene, putative calcium binding protein, mouse embryo.
MO25 was isolated from the 8-cell stage of mouse embryos by differential screening, its level being increased relative to unfertilised eggs.[50] Function of the gene product is unknown and it is dissimilar to other known proteins, except the C-terminus of the predicted protein has similarity to the calcium binding site of EF-hand proteins, such as a human plasma membrane calcium pump (ATPase).

349 Blast 152 (3.4e-15) S P15170: GST1-HS GTP-binding protein, human.
Also homologous to sup2 of yeasts. These are elongation factor 1α

homologues, involved in protein synthesis (EF1 mediates interaction between the ribosome and aminoacyl-tRNA) and in the G1-to-S phase transition in yeast[51] and regulation of mammalian cell growth.[52] Sup2 is essential for viability. Match is in the C-terminus which is well conserved among the mammalian and yeast proteins. Involvement in protein synthesis and cell growth logical for the castor endosperm or embryo used to make the library, which are actively growing and synthesizing storage proteins. Note also: CRS680 is elongation factor 2.

401 Blast 89 (2.7e-14) same as for pCRS349, above.

438 Blast 110 (5.0e-20) G X14977: aldehyde dehydrogenase, rat.
Also match to the human homologue, which are both mitochondrial, but matches to cytosolic forms are also observed. [e.g. acetaldehyde + NAD^+ + H_2O → acetic acid + NADH]

455 Blast 141 (3.7e-12) P S22439: NAM8 protein, yeast.
Encodes a protein with putative RNA binding motifs and acts as a supressor of mitochondrial splicing deficiencies when overexpressed. Also a match for NGR1, negative growth regulatory protein, of yeast, which encodes a putative glucose repressible protein containing two RNA recognition motifs. There is weaker similarity (in the same reading frame) to other RNA binding proteins.

495 Blast 154 (3.8e-14) S P23913: lamin B receptor, chicken.
Also 145 (7.2e-13) G S49653: sterol C-14 reductase, *Saccharomyces cerevisiae*; 90 (6.5e-05) S P25340: hypothetical transport protein YGL022, *Saccharomyces cerevisiae*; and 84 (0.00047) P A43765: sts1+ protein, *Schizosaccharomyces pombe*. The match of the castor sequence to these four apparently disparate sequences is at the same conserved residues, suggesting that they all have something in common. The lamin receptor is on the inner membrane of the nuclear envelope. Alignment of castor sequence and lamin receptor includes parts of two potential transmembrane domains. The sterol reductase is quite likely a membrane protein since its substrate is membrane-soluble, but the cellular location of the enzyme has not been described.[53] Alignment with the sterol reductase involves matches between similar residues as for the lamin receptor. Alignment of castor sequence and YGL022 (which is a different protein to sterol reductase) also includes part of one and the whole of another potential transmembrane domain, but sequence similarity is sparse (25/95 identities, 42/95 positives). $Sts1^+$, a gene of unknown function and pleiotropic effects when mutated has been suggested to be an integral membrane protein on the basis of sequence similarity between $sts1^+$, the lamin B receptor and YGL022,[54] all three proteins having a similar hydrophobicity pattern consisting of eight or nine putative transmembrane domains. It seems that these four characterised genes and CRS495 may consitute a class of

similar membrane proteins with differing functions.

545 Blast 122 (1.5e-10) P S25013: vacuolar ATPase 14 kDa chain, *Manduca sexta* (tobacco hornworm).
This is the only match found. The *Manduca* ATPase is unpublished. See also CRS785, which is a clone from the same gene, and CRS512, which is a match for the vacuolar H^+ ATPase catalytic subunit of cotton, with a weak alignment also to chain A of the *Manduca* ATPase. Rapid enlargement of the protein-storing vacuoles in the developing castor seed would require the concomitant synthesis of vacuole-associated proteins, such as the ATPase (which furthermore may be involved in protein import), and tonoplast intrinsic protein (see CRS279, Table 11).

548 Blast 413 (2.4e-56) G X56932: 23 kDa highly basic protein, human.
Also 411 (4.7e-56) S P19253: transplantation antigen P198, mouse. P198 is also highly basic. The function of P198 is not known.[55]

592 Blast 100 (7.1e-12) P S27783: hypothetical protein 1 (adenylate cyclase homologue), *C. elegans*.

633 Blast 186 (4.7e-20) P S27951: renal cortical Na/Pi cotransporter, rabbit.
A collaboration has been initiated with Dr. Yves Poirier, who has isolated and sequenced two different cDNAs of *Arabidopsis* homologues, using pCRS633 as a probe, which are found to encode proteins having the characteristic transmembrane domains of transporters.[56] Preliminary data suggests that neither of these cDNAs maps to the chromosomal location of the *pho1* mutation involved in phosphate translocation to the shoot.[57] Few genes directly involved in nutrient uptake in plants have been cloned, and none involved in phosphate nutrition. This clone may be a useful beginning.

636 Blast 132 (7.8e-12) S P13641: autoantigen small nuclear ribonuleoprotein SM-D, human.
Also a very similar match for the mouse homologue. Nucleic acid binding protein involved in mRNA splicing. SM-D is one of the core proteins of the small nuclear RNP complexes, suggesting that the castor clone is probably for a nuclear-localised ribonucleoprotein. Both human and mouse proteins have a lysine rich domain and Gly-Arg repeats at carboxy terminus, but their functional significance is unclear.[58] These are also found in the castor sequence, and do not directly resemble any known plant or animal nuclear localisation sequences.[59]

666 Blast 442 (6.6e-61) G Z19599: epsilon isoform of 14-3-3 protein, mouse.
Similar matches to homologues from many other organisms including plants, but the castor sequence is more similar to the mouse, sheep and yeast proteins

than the homologue from barley, suggesting that this isoform is not previously sequenced from higher plants. The mammalian proteins are protein kinase II and protein kinase C inhibitors. The barley homologue is induced upon attempted penetration by powdery mildew fungi.

669 Blast 171 (2.8e-17) P A37863: ferrochelatase (haem synthetase), mouse.
Also similar matches to other mammals, bacteria, and yeast, but this is presumably the first sequence from plants. This mitochondrial enzyme catalyses the last step in haem biosynthesis, in which protoporphyrin IX binds Fe^{2+}. A collaboration has been initiated with Dr. Mary Lou Guerinot, Dartmouth College.

680 Blast 502 (4.8e-65) S P28996: elongation factor 2, *Chlorella kessleri*.
As well as this lower-plant match, there are matches to elongation factor 2 of other eukaryotes, and elongation factor G of prokaryotes, but this appears to be the first higher plant sequence. EF-2/EF-G is required for the GTP-hydrolysing translocation of the ribosome from one codon to the next, during which the nascent peptide chain is switched from the A to the P site and the uncharged tRNA is released. This, along with the ribosomal proteins, is an unsurprising sequence to find expressed in the developing castor seed where seed storage protein synthesis is active. EF-2 (but not EF-G) is the target for inactivation by diptheria toxin (by ADP-ribosylation of the novel amino acid dipthamide) and ricin.[60] Note also that CRS349, CRS401 are elongation factor 1α homologues.

692 Blast 89 (6.5e-15) S Q01888: mitochondrial solute carrier protein homologue, bovine.
Similar score mitochondrial ATP/ADP carrier protein, human, and many homologues from other organisms. The human ATP/ADP carrier protein is in mitochondrial inner membrane. Maize brittle-1 appears to be the highest-scoring plant match, and may be an amyloplast homologue of the mitochondrial carrier proteins, found in endosperm of developing kernels. *Brittle-1* mutants have severely reduced amounts of starch deposition in the endosperm. Sequence analysis of the *Brittle-1* gene revealed a plastid transit peptide, and it was suggested that the Brittle-1 protein may be an amyloplast envelope translocator, important for ATP import for starch synthesis.[61] Could this be a leukoplast version in castor? Uptake of ATP by the plastids of developing castor endosperm may be necessary for fatty acid synthesis (see CRS304), as it is necessary for starch synthesis in maize. CRS523, CRS581, and CRS630 appear to be quite different ADP/ATP carrier proteins, very strongly homologous to other such proteins from higher plants (CRS523 giving a score of 584 with a potato protein, and only 82 with brittle-1, CRS581 a score of 348 with an *Arabidopsis* protein and 70 with brittle-1, CRS630: 474 for rice, 70 for brittle-1).

729 Blast 162 (1.9e-18) G X15141: histone H4, *Physarum polycephalum* (slime mould).
Many very similar matches for other histone H4 genes, including wheat. Very high conservation that is characteristic of histones: 32/38 identical residues with the *Physarum* protein. Note: CRS808 is an histone H2A homologue.

741 Blast 224 (3.6e-30) S P13641: autoantigen small nuclear ribonucleoprotein SM-D, human.
Also a very similar match for the mouse homologue. These are the same proteins which match CRS636, but this sequence is substantially different to CRS636 (approx 40% identity at the nucleotide level, and approx 25% at the amino acid level), suggesting that this is a second, quite different homologue of SM-D. CRS636 has 27/86 (31%) identical amino acids to human SM-D, while CRS741 has 45/68 (66%) identical to SM-D. CRS741 also has the conserved lysines and Gly-Arg repeats, and also shows matches to the Gly-Arg repeats of another nuclear protein, of human herpesvirus.

753 Blast 93 (1.3e-11) S P21187: poly(A) binding protein, *Drosophila melanogaster.*
Also matches yeast poly(A) binding protein, human initiation factor eIF-4B, as well as weaker matches to other higher plant RNA-binding proteins, e.g. *Nicotiana, Arabidopsis*. The castor sequence is most similar to the *Drosophila* protein in a short domain identified to be involved in RNA binding (5 identical residues of 8 (63%) compared to 19/66 (28%) overall).

763 Blast 94 (7.4e-14) S P19623: spermidine synthase (putrescine aminopropyltransferase), human.
Last step in the biosynthesis of spermidine from arginine and methionine: SAMamine + putrescine → methylthioadenosine + spermidine, i.e. transfers amino propyl group (H_2N-[CH_2]$_3$). Also a weaker match to the same enzyme from *E. coli*, and also a much weaker match in the same reading frame to aspartate aminotransferase of *Sulfulobus solfataricus*. Other enzymes involved in polyamine biosynthesis were also identified. These are enzymes of arginine biosynthesis (CRS519 and CRS772, see Table 8), *S*-adenosylmethionine synthetase (CRS554), and *S*-adenosylmethionine decarboxylase (pCRS555). Polyamine metabolism is discussed further in the text.

785 Blast 275 (3.3e-34) P S25013: vacuolar ATPase 14 kDa chain, *Manduca sexta* (tobacco hornworm).
Match to the same protein as CRS545, except CRS785 is from a much longer clone. There are few nucleotide sequence differences between CRS545 and CRS785 in the region of overlap, probably ascribable to sequencing errors, thus CRS545 and CRS785 probably represent clones originating from the same gene. Again, this was the only match in the databases.

865	Blast 144 (5.4e-13) S P12613: T-complex protein 1 homologue, *Drosophila melanogaster*. Cytosolic molecular chaperone, involved in folding of actin and tubulin. An *Arabidopsis* homologue has been cloned, but is a slightly weaker match to the castor sequence than are the mammalian and yeast proteins, suggesting that this may be a new isoform for plants.

Table 10. Clones homologous to ribosomal proteins. Those having homology to ribosomal proteins for which higher plant sequences are already available, and those having homology to ribosomal proteins apparently previously unsequenced from higher plants (see text), are listed separately. The clones are listed by the suffix of their clone number (pCRS416 etc.). Information for each clone is structured as follows: blastx score, (probability of random alignment), database (G: GenBank; S: SwissProt; P: PIR), accession number, protein description, organism.

From Higher Plants

416 Blast 242 (6.1e-30) G Z17784: 40S ribosomal protein S13, *Arabidopsis thaliana.*

432 Blast 373 (4.1e-51) S P19950: 40S ribosomal protein S14, *Zea mays.*

497 Blast 275 (7.2e-36) P S25550: ribosomal protein S6, *Nicotiana tabacum.*

446 Blast 412 (4.5e-53) P S21519: acidic ribosomal protein P0 homologue, *Chenopodium rubrum.*

546 Blast 466 (3.9e-62) P S19893: cytosolic ribosomal protein L8, *Lycopersicon esculentum.*

571 Blast 86 (6.7e-14) P S25983: mitochondrial ribosomal protein S11, *Marchantia polymorpha.*

595 Blast 299 (6.9e-40) G Z17767: 60S ribosomal protein L27A, *Arabidopsis thaliana.*

684 Blast 348 (2.4e-44) S P17094: 60S ribosomal protein L3, *Arabidopsis thaliana.*

707 Blast 263 (4.0e-34) S Q00332: ribosomal protein S15A, *Brassica napus.*

710 Blast 227 (8.2e-28) S P29766: 60S ribosomal protein L8, *Lycopersicon esculentum.*

766 Blast 508 (7.3e-68) S P17094: 60S ribosomal protein L3, *Arabidopsis thaliana.*

837 Blast 452 (5.7e-62) S Q00332: ribosomal protein S15A, *Brassica napus.*

From Other Eukaryotes

262 Blast 230 (5.0e-28) S P08636: 40S ribosomal protein S17, chicken.

312 Blast 476 (4.8e-65) G M96570: ribosomal protein S28, *Saccharomyces cerevisiae.*

346 Blast 303 (1.0e-37) S P19889: acidic ribosomal protein PO/DNAse, *Drosophila melanogaster.*

356 Blast 279 (3.4e-34) S P02362: 40S ribosomal protein S8, *Xenopus laevis.*

358 Blast 424 (6.2e-56) P S25633: ribosomal YL10 protein homologue, *Chironomus tentans.*

377 Blast 266 (3.4e-33) G X68202: ribosomal protein L27, *Chlamydobotrys*

stellata.

407 Blast 102 (4.1e-09) S P29314: 40S ribosomal protein S9, rat.

409 Blast 139 (3.7e-12) G M77233: ribosomal protein, human.

426 Blast 342 (7.8e-46) G X68202: ribosomal protein L27, *Chlamydobotrys stellata*.

442 Blast 276 (8.4e-36) G X68202: ribosomal protein L27, *Chlamydobotrys stellata*.

522 Blast 122 (2.0e-22) S P05387: 60S acidic ribosomal protein P2, human.

643 Blast 178 (1.9e-19) S P18124: 60S ribosomal protein L7, human.

655 Blast 248 (8.7e-29) S P26321: 60S ribosomal protein L5, *Saccharomyces cerevisiae*.

683 Blast 81 (0.00029) G M76718: 60S ribosomal protein rpl29, *Tetrahymena thermophila*.

690 Blast 295 (4.5e-39) P S24989: ribosomal protein L31, *Chlamydomonas reinhardtii*.

701 Blast 170 (2.5e-20) G M96570: ribosomal protein S28, *Saccharomyces cerevisiae* .

708 Blast 267 (3.4e-34) S P05744: 60S ribosomal protein YL37, *Saccharomyces cerevisiae*.

709 Blast 221 (4.2e-26) S P05387: 60S acidic ribosomal protein P2, human.

744 Blast 367 (4.9e-49) S P25121: 60S ribosomal protein L11, rat.

814 Blast 80 (0.00030) G Z21487: ribosomal protein L21, *Saccharomyces cerevisiae*.

825 Blast 258 (8.1e-33) S P25111: 40S ribosomal protein S25, human.

860 Blast 107 (5.8e-10) S P24050: 40S ribosomal protein S5, rat.

Table 11. Clones for which the most similar match in the databases was from a higher plant, other than ribosomal proteins. Clones are grouped according to the general area of metabolism or cell function with which they may be involved (see text for discussion). The clones are listed by the suffix of their clone number (pCRS792 etc.). Information for each clone is structured as follows: blastx score, (probability of random alignment), database (G: GenBank; S: SwissProt; P: PIR), accession number, protein description, organism.

Lipid Synthesis

792 Blast 105 (5.1e-08) G Z12823: fructokinase, *Solanum tuberosum.*
517 Blast 501 (7.0e-66) S P21343: 6-phoshofructokinase, *Solanum tuberosum.*
427 Blast 531 (3.7e-68) S P08440: fructose-bisphosphate aldolase, *Zea mays.*
823 Blast 572 (3.8e-76) G D13512: cytoplasmic aldolase, *Oryza sativa.*
294 Blast 208 (1.3e-24) G X59517: glycolytic glyceraldehyde-3-phosphate dehydrogenase, *Antirrhinum majus.*
330 Blast 420 (5.8e-53) P JQ1186: enolase, *Lycopersicon esculentum.*
380 Blast 573 (2.4e-76) P JQ1186: enolase, *Lycopersicon esculentum.*
415 Blast 346 (3.2e-44) P JQ1186: enolase, *Lycopersicon esculentum.*
439 Blast 371 (2.5e-46) P JQ1186: enolase, *Lycopersicon esculentum.*
509 Blast 561 (2.2e-73) P JQ1186: enolase, *Lycopersicon esculentum.*
503 Blast 354 (7.8e-47) G X17362: mitochondrial malate dehydrogenase, *Citrullus vulgaris.*
504 Blast 127 (8.4e-11) P S18152: dihydrolipoamide dehydrogenase precursor, *Pisum sativum.*
844 Blast 153 (2.6e-34) P S22384: dihydrolipoamide dehydrogenase, *Pisum sativum.*
544 Blast 317 (6.1e-42) S P08817: acyl carrier protein II precursor, *Hordeum vulgare.*
601 Blast 176 (2.3e-22) S P07854: acyl carrier protein I, *Spinacia oleracea.*
507 Blast 554 (7.1e-75) G L13242: beta-ketoacyl-ACP synthase, *Ricinus communis.*
794 Blast 155 (1.4e-16) S P28643: 3-oxoacyl-ACP reductase precursor, *Cuphea lanceolata.*
291 Blast 292 (2.7e-38) S P22243: stearoyl-ACP desaturase, *Carthamus tinctorius.*
667 Blast 255 (2.8e-31) G M91238: stearoyl ACP desaturase, *Solanum tuberosum.*
858 Blast 163 (1.9e-17) S P22243: stearoyl-ACP desaturase, *Carthamus tinctorius.*
677 BLAST 110 (2.7E-16) G D14410: ORF, *VIGNA RADIATA*; BLAST 112 (3.2E-13) G L01418: LINOLEIC ACID DESATURASE, *BRASSICA NAPUS.*
834 Blast 126 (6.8e-21) P A44227: OMEGA-3 FATTY ACID DESATURASE, *BRASSICA NAPUS*; 131 (2.7e-20) G D14410: ORF, *VIGNA RADIATA.*
698 Blast 134 (1.3e-12) G M87514: cytochrome b5, *Brassica oleracea.*

Table

384
398

293

494

728

Pr

29

8
8
2

384 Blast 120 (1.2e-10) S P29111: major oleosin nap-II, *Brassica napus*.
398 Blast 256 (2.4e-29) P JQ0986: lipid body-associated membrane protein, *Daucus carota*.
293 Blast 482 (2.0e-66) G S70711: protein disulfide isomerase homologue, *Medicago sativa*.
494 Blast 517 (4.9e-67) P S22479: protein disulphide isomerase precursor, *Medicago sativa*.
728 Blast 384 (9.2e-51) S P29828: protein disulphide isomerase precursor, *Medicago sativa*.

Protein Synthesis

299 Blast 222 (3.5e-27) G M92353: anthranilate synthase alpha subunit, *Arabidopsis thaliana*.
841 Blast 267 (1.4e-33) P S28429: alanine transaminase, *Panicum miliaceum*.
856 Blast 406 (9.4e-55) G L09702: aspartate aminotransferase, *Glycine max*.
267 Blast 199 (1.2e-22) G X59802: cruciferin, *Raphanus sativus*. Shouldn't have been sequenced - seed positive.
269 Blast 130 (4.8e-12) S P09800: legumin precursor, *Gossypium hirsutum*.
298 Blast 326 (9.5e-42) S P09800: legumin precursor, *Gossypium hirsutum*.
404 Blast 259 (2.7e-31) S P09800: legumin precursor, *Gossypium hirsutum*.
405 Blast 105 (3.5e-08) S P09800: legumin precursor, *Gossypium hirsutum*.
408 Blast 169 (6.0e-19) S P09800: legumin precursor, *Gossypium hirsutum*.
434 Blast 134 (1.4e-13) G X59802: cruciferin, *Raphanus sativus*.
443 Blast 99 (1.1e-06) S P09800: legumin precursor, *Gossypium hirsutum*.
453 Blast 276 (1.1e-32) S P09800: legumin precursor, *Gossypium hirsutum*.
454 Blast 182 (1.8e-19) P A35540: cruciferin precursor, *Brassica napus*.
520 Blast 154 (1.2e-15) G X14393: preproglutelin, *Oryza sativa*.
540 Blast 163 (6.2e-20) G X59804: cruciferin, *Raphanus sativus*.
584 Blast 400 (3.5e-52) S P13744: 11S globulin beta subunit precursor, *Cucurbita pepo*.
641 Blast 215 (4.4e-28) G X57849: cruciferin cru2/3 subunit, *Brassica napus*.
686 Blast 343 (1.4e-44) G X59807: cruciferin, *Raphanus sativus*.
691 Blast 253 (3.3e-32) G X14312: CRA1 12S seed storage protein, *Arabidopsis thaliana*.
739 Blast 252 (1.2e-32) G X14312: CRA1 12S seed storage protein, *Arabidopsis thaliana*.
811 Blast 117 (5.9e-10) S P09800: legumin precursor, *Gossypium hirsutum*.
813 Blast 372 (2.1e-46) S P09800: legumin precursor, *Gossypium hirsutum*.
281 Blast 183 (1.4e-20) G X54158: 2S albumin, *Ricinus communis*.
328 Blast 183 (2.2e-20) G X54158: 2S albumin, *Ricinus communis*.
337 Blast 113 (2.9e-18) G X54158: 2S albumin, *Ricinus communis*.
362 Blast 245 (3.1e-32) G X54158: 2S albumin, *Ricinus communis*.
375 Blast 521 (2.7e-70) S P01089: 2S albumin precursor, *Ricinus communis*.

431 Blast 319 (1.4e-39) G X54158: 2S albumin, *Ricinus communis*.
549 Blast 82 (3.7e-05) G X54158: 2S albumin, *Ricinus communis*.
735 Blast 96 (2.8e-08) P B25802: 2S seed storage protein large chain, *Bertholletia excelsia*.
780 Blast 78 (1.4e-05) G X54158: 2S albumin, *Ricinus communis*.
807 Blast 339 (5.8e-50) G X54158: 2S albumin, *Ricinus communis*.
816 Blast 243 (7.0e-31) G X54158: 2S albumin, *Ricinus communis*.
424 Blast 440 (1.9e-60) P A24010: ricin D chain B, *Ricinus communis*.
467 Blast 200 (9.4e-22) P A24210: agglutinin B chain, *Ricinus communis*.
826 Blast 570 (1.4e-75) P A24210: agglutinin chain B, *Ricinus communis*.
385 Blast 188 (4.3e-21) S P24922: initiation factor 5A, *Nicotiana plumbaginifolia*.
359 Blast 93 (3.2e-05) G L00623: opaque2 heterodimerizing protein, *Zea mays*.
264 Blast 324 (1.6e-42) S P29357: chloroplast membrane 70kDa heat shock protein, *Spinacia oleracea*.
348 Blast 245 (7.2e-29) P S25005: heat shock protein 70kDa, *Phaseolus vulgaris*.
382 Blast 544 (2.1e-72) P JQ1360: luminal binding protein BLP-4 precursor, *Nicotiana tabacum*.
397 Blast 363 (8.4e-48) P JS0710: low molecular weight heat shock protein, *Oryza sativa*.
598 Blast 407 (7.4e-54) G M87646: chaperonin 10, *Spinacia oleracea*.
657 Blast 206 (2.1e-23) P S25005: heat shock protein 70K, *Phaseolus vulgaris*.
759 Blast 382 (6.5e-50) P JS0710: low molecular weight heatr shock protein, *Oryza sativa*.
778 Blast 324 (5.7e-42) G X67695: DNAJ-1 (mitochondrial heat shock protein), *Cucumis sativus*.
820 Blast 295 (2.1e-37) S P29357: chloroplast envelope membrane 70 KD heat shock-related protein, *Spinacia oleracea*.
833 Blast 211 (1.3e-26) S P27879: 18.1 KD class 1 heat shock protein, *Medicago sativa*.
275 Blast 255 (1.4e-30) G M95796: St12p protein, *Arabidopsis thaliana*.
279 Blast 86 (1.6e-05) G S52690: nodulin-26 homologue, *Pisum sativum*.
360 Blast 81 (9.2e-05) G S52690: nodulin-26 homologue, *Pisum sativum*.
663 Blast 517 (3.7e-72) S P26587: tonoplast intrinsic protein alpha, *Arabidopsis thaliana*.
749 Blast 230 (1.3e-28) S P23958: tonoplast intrinsic protein alpha, *Phaseolus vulgaris*.
765 Blast 438 (2.2e-59) P JG1106: tonoplast intrinsic protein alpha, *Phaseolus vulgaris*.
831 Blast 396 (3.5e-53) P JQ1106: tonoplast intrinsic protein alpha, *Phaseolus vulgaris*.
512 Blast 279 (6.0e-35) G L03186: vacuolar H^+-ATPase catalytic subunit, *Gossypium hirsutum*.

Ricinine Biosynthesis

696 Blast 214 (3.7e-25) S Q00763: bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase, *Populus tremuloides*.
738 Blast 140 (1.6e-12) G X72593: naringenin, 2-oxoglutarate 3-dioxygenase, *Callistephus chinensis*.

Polyamine Biosynthesis

554 Blast 458 (2.3e-62) G M62758: S-adenosylmethionine synthetase, *Petroselinum crispum*.
555 Blast 136 (2.2e-12) P S28047: TUB13 protein, *Solanum tuberosum*.

Organ-Specific Expression

371 Blast 80 (0.001) G Z18891: BP8 gene product, *Betula pendula*.
565 Blast 120 (3.3e-09) S P21746: embryonic abundant protein precursor, *Vicia faba*.
589 Blast 301 (5.2e-39) P JQ1107: 18.3k protein precursor, pollen, *Zea mays*.
406 Blast 172 (8.5e-18) S P16148: pplzo2 protein, *Lupinus polyphyllus*.
430 Blast 131 (1.8e-11) S P16146: pplzo2 protein, *Lupinus polyphyllus*.

Defense-Related Proteins

413 Blast 180 (8.1e-23) G X69139: protease inhibitor II, *Arabidopsis thaliana*.
648 Blast 200 (4.6e-26) G X69139: protease inhibitor II, *Arabidopsis thaliana*.
653 Blast 198 (1.5e-26) G X69139: protease inhibitor II, *Arabidopsis thaliana*.
754 Blast 215 (2.7e-29) P S24965: p322 protease inhibitor, *Glycine max*.
580 Blast 137 (2.4e-13) P G28027: protein P10, *Nicotiana* sp.
711 Blast 109 (3.0e-08) S P28493: thaumatin-like protein precursor, *Arabidopsis thaliana*.

Dessication-Related Proteins

521 Blast 137 (1.8e-13) P S25121: dehydrin-cognate, *Pisum sativum*.
557 Blast 245 (8.2e-28) S P22242: dessication-related protein clone PCC13-62 precursor, *Craterostigma plantagineum*.

Other

Protein Kinases
280 Blast 483 (6.2e-67) G D10152: protein tyrosine-serine-threonine kinase, *Arabidopsis thaliana*.

436 Blast 231 (1.4e-26) G L08789: protein kinase, *Arabidopsis thaliana.*
731 Blast 428 (9.6e-57) G L05562: protein kinase, *Arabidopsis thaliana.*
733 Blast 550 (2.2e-73) P S27760: protein kinase 6, *Glycine max.*
817 Blast 308 (1.1e-38) S P28583: calcium-dependent protein kinase, *Glycine max.*
836 Blast 509 (2.6e-67) P S26627: serine-protein kinase, *Arabidopsis thaliana.*

ADP/ATP Carrier Proteins
581 Blast 348 (3.4e-47) G X68592: adenosine nucleotide translocator, *Arabidopsis thaliana.*
523 Blast 584 (2.6e-79) P S17917: ADP, ATP carrier protein precursor, *Solanum tuberosum.*
630 Blast 474 (2.6e-63) P JS0711: ATP/ADP translocator protein, *Oryza sativa.*

Ubiquitin-Related Proteins
342 Blast 382 (3.9e-53) G L06967: ubiquitin carrier protein, *Medicago sativa.*
505 Blast 564 (3.5e-76) G M74100: ubiquitin fusion protein, *Nicotiana sylvestris.*
387 Blast 343 (4.1e-47) G M74100: ubiquitin fusion protein, *Nicotiana sylvestris.*
672 Blast 236 (9.5e-33) G LO6967: ubiquitin carrier protein, *Medicago sativa.*

Actin Depolymerizing Factor
313 Blast 158 (4.8e-16) P S25061: actin depolymerizing factor, *Lilium longiflorum.*
363 Blast 212 (7.3e-25) P S25059: actin depolymerizing factor, *Brassica napus.*
524 Blast 320 (4.9e-44) P S25061: actin depolymerizing factor, *Lilium longiflorum.*

Tubulin
558 Blast 578 (6.8e-78) P S23221: tubulin alpha chain, *Prunus amygdalus.*
781 Blast 626 (5.7e-85) P S23221: tubulin alpha chain, *Prunus amygdalus.*

Ras/G-Proteins
322 Blast 502 (9.3e-69) G S76123: ras-related ypt family cDNA yptm2, *Zea mays.*
343 Blast 344 (3.7e-46) P S25543: RAS-related GTP-binding protein, *Pisum sativum.*

Metallothionein
444 Blast 423 (6.3e-60) G L02306: RCMT1 metallothionein, *Ricinus communis.*
746 Blast 99 (1.3e-07) S P30564: metallothionein I homologue, *Ricinus communis.*

Miscellaneous
305 Blast 249 (3.9e-51) P S22499: hypothetical protein, *Bromus secalinas.*
743 Blast 171 (7.0e-19) P S22499: hypothetical protein, *Bromus secalinas.*
447 Blast 150 (7.1e-20) G X60391: proline-rich protein, *Phaseolus vulgaris.*
499 Blast 242 (1.7e-32) S P10973: non-specific lipid transfer protein A, *Ricinus communis.*
511 Blast 466 (1.9e-62) S P12412: legumain endopeptidase, *Vigna mungo.*

596 Blast 89 (8.0e-06) P S19253: gene le25 protein, *Lycopersicon esculentum.*
700 Blast 522 (4.0e-72) P JQ0939: cyc07 protein, S-phase specific, *Catharanthus roseus.*
720 Blast 179 (1.5e-19) S P19595: UTP-glucose-1-phosphate uridylyltransferase, *Solanum tuberosum.*
779 Blast 256 (3.2e-34) S P28756: superoxide dismutase (Cu-Zn), *Oryza sativa.*
805 Blast 557 (2.7e-77) G M90504: RNA polymerase II, *Glycine max.*
808 Blast 312 (2.2e-41) G X67819: histone H2A, *Picea abies.*
818 Blast 134 (4.0e-12) G L20864: ascorbate peroxidase, *Spinacia oleracea.*
842 Blast 148 (1.1e-17) S P13089: aux28 protein, *Glycine max.*

Lipid Synthesis

Lipid synthesis is a dominant area of metabolism in the developing castor seed, such that at maturity, oil constitutes 50% of the seed dry weight. Not surprisingly therefore, many genes with a likely or certain role in some aspect of lipid synthesis were identified.

We begin by considering the source of carbon for fatty acid and lipid synthesis. Sucrose imported by the seed is inverted to glucose and fructose, which enter the glycolysis pathway. Both are converted to their 6-phosphates; **pCRS792** encodes fructokinase. Glucose-6-phosphate is isomerised to fructose-6-phosphate, and fructose-6-phosphate from both sources is phosphorylated to fructose-1,6-diphosphate by 6-phosphofructokinase (**pCRS517**). Fructose-1,6-diphosphate is then cleaved to dihydroxyacetone phosphate and 3-phosphoglyceraldehyde by aldolase (**pCRS427**, **pCRS823**).

At this point we are interested in a branch of the pathway, in which dihydroxyacetone phosphate (directly from the aldolase reaction or isomerized from 3-phosphoglyceraldehyde) is converted to glycerol-3-phosphate by glycerol-3-phosphate dehydrogenase (**pCRS294**). Glycerol-3-phosphate is the carbon backbone for the synthesis of glycerolipids, including the major storage lipid, triacylglycerol.

Several steps further in the glycolysis pathway, 2-phosphoglyceric acid is dehydrated to phosphoenol pyruvate by enolase (**pCRS330**, **pCRS380**, **pCRS415**, **pCRS439** and **pCRS509**; see also Table 7). Phosphoenol pyruvate is converted to enol pyruvate (and thus spontaneously to pyruvate) by pyruvate kinase (**pCRS304**,

Table 9), completing glycolysis.

An alternative source of pyruvate has been proposed to be active in the developing castor seed.[62] Rather than being converted to pyruvate by pyruvate kinase, phosphoenol pyruvate would be carboxylated to oxaloacetate in the cytoplasm, and then converted to malate by malate dehydrogenase (**pCRS503**). Malate is then imported by the plastid and converted to pyruvate by malic enzyme.

Pyruvate, derived from either source, is converted to acetyl-CoA by the pyruvate dehydrogenase complex (PDC). The PDC has three subunits: E1 (**pCRS798**, Table 8), dihydrolipoamide acetyltransferase, and dihydrolipoamide dehydrogenase (**pCRS504**, **pCRS844**; these two clones share almost no similarity at the nucleotide level, and probably represent the plastid and mitochondrial forms, respectively, since pCRS844 has much higher amino acid identity (86%) with the previously-sequenced pea mitochondrial protein than does pCRS504 (45%)). A collaboration has been initiated Dr Christoph Benning, IGF Berlin, who will use the three PDC clones to characterise the genes of plastid PDC, and their possible manipulation for increasing seed oil production.

Acetyl-CoA is carboxylated to malonyl-CoA, which is used for fatty acid synthesis. This is accomplished by acetyl-CoA carboxylase, of which **pCRS706** (see Table 8) may encode a subunit. Fatty acid synthesis occurs in the plastid with the growing acyl chain esterified to acyl carrier protein (ACP). Two ACP clones were identified, **pCRS544** and **pCRS601**, which appear to be independent clones derived from the same gene, since they show 100% nucleotide identity in the region of

overlap. The fatty acid synthase cycle involves condensation of the acyl-ACP with malonyl-ACP, catalysed by 3-ketoacyl-ACP synthase (**pCRS507**, which shows 100% identity to a gene previously sequenced from castor, but which has not been assigned to one of the three possible isozymes), reduction to 3-hydroxyacyl-ACP by 3-ketoacyl-ACP reductase (**pCRS794**), and dehydration to enoyl-ACP by 3-hydroxyacyl-ACP dehydratase (the putative identity of **pCRS673**; see Table 8). The cycle is completed by enoyl-ACP reductase.

In the castor seed, approximately 90% of the fatty acid synthesized is modified by the action of two enzymes, producing ricinoleic acid. The first is the soluble stearoyl-ACP desaturase (**pCRS291**, **pCRS667**, and **pCRS858**) which introduces a double bond between carbons 9 and 10 (counting from the carboxyl end). The second enzyme, oleate-12-hydroxylase, is the principle enzyme for which a clone was sought in this project. As described above, it was hypothesized that this gene is homologous to those of the microsomal fatty acid desaturases, and it was hoped that the oleate-12-hydroxylase gene could be identified on this basis. Two clones, **pCRS677** and **pCRS834**, gave such a match to the *fad3* gene, encoding the microsomal ω-3 desaturase, the only known higher plant membrane-bound desaturase which was in the databases. These two clones were therefore considered putative oleate-12-hydroxylase clones and their analysis is the subject of chapter 6.

In addition, electron transfer to the microsomal desaturases and oleate-12-hydroxylase requires cytochrome b_5 (chapter 2), and a cytochrome b_5 clone (**pCRS698**) was also identified.

Fatty acids are sequentially esterified to glycerol-3-phosphate to form the storage lipid, triacylglycerol. Triacylglycerol accumulates in the cell in spherosomes (oil bodies), which are bounded by a phospholipid monolayer studded with characteristic proteins, oleosins (**pCRS384, pCRS398**; these are independent clones, but further sequence data is required to determine whether they represent the same or different genes). The mechanism of transport of triacylglycerol from the site of synthesis in the endoplasmic reticulum to the spherosome is unknown. In this light, it is interesting to note that protein disulphide isomerase (PDI; **pCRS293, pCRS494,** and **pCRS728**) is an essential subunit of the mammalian liver triacylglycerol-specific microsomal lipid transfer protein (MTP).[63] In liver, MTP may mediate the transport of triacylglycerol to nascent very low density lipoprotein particles in the lumen of the endoplasmic reticulum and golgi aparatus. The other subunit of MTP has no activity when separated from PDI, and has no homology to other known proteins.[64] A role for PDI in triacylglycerol transfer in the developing castor seed is purely speculative, and the abundance of PDI clones can also be rationalised by its known role in protein folding (below).

Protein Synthesis

In addition to storing lipid, the castor seed accumulates storage proteins, and these are synthesized at similar stages of seed development to those selected (on the basis of active lipid synthesis) for construction of the cDNA library. All aspects of storage protein synthesis are considered here, from nitrogen import into the

developing seed, to storage of the synthesized protein in the vacuole.

Import of nitrogen by the developing seed must be co-ordinated with its assimilation into amino acids, and this may suggest a role for a nitrogen regulatory protein PII homologue (**pCRS852**, see Table 8).

Various clones identified may encode enzymes of amino acid biosynthesis. These are threonine synthase (**pCRS265**, see Table 8); anthranilate synthase, involved in tryptophan biosynthesis (**pCRS299**); two enzymes of arginine biosynthesis: carbamoyl-phosphate synthase and ornithine carbamoyltransferase (**pCRS519** and **pCRS772** respectively, see Table 8); and three aminotransferases: alanine transaminase (**pCRS841**), aspartate aminotransferase (**pCRS856**), and 4-aminobutyrate aminotransferase (**pCRS336**, see Table 8).

The most abundant transcripts in the developing seed are probably those encoding the seed storage proteins themselves. Attempts were made, by differential screening, to eliminate these from the pool of clones to be sequenced. This was largely successful, but some such clones were sequenced nevertheless, and are listed in Table 11. Short clones gave a weaker hybridisation signal, making them more difficult to resolve from the less abundant clones which targetted for sequencing. In addition to the 2S and 12S seed storage proteins, the castor seed accumulates the toxic protein ricin, to about 1.5% of the total protein content. Several clones (**pCRS424**, **pCRS467**, and **pCRS826**) encoding ricin or the closely related agglutinin were sequenced.

A large number of clones encoding a great variety of ribosomal proteins were

sequenced, and are listed in Table 10. These include those previosly sequenced from higher plants, and those for which the strongest match was not to a higher plant gene, and therefore have presumably not previously been sequenced from higher plants. It should be noted that the databases used for clone identification in this project do not include dbEST, in which expressed sequence tags (ESTs) from projects similar to this one are deposited. Ribosomal protein genes are also a prominent class of clones obtained in the other plant EST projects, and there may be some overlap between those identified in this and the other projects.

A number of clones were identified that encode components of protein synthesis other than ribosomal proteins. These components are: asparaginyl-tRNA synthetase (**pCRS296**, see Table 8); initiation factor 5A (**pCRS385**); elongation factor 1α (**pCRS349** and **pCRS401**, see Table 9); and elongation factor 2 (**pCRS680**, see Table 9).

Some proteins are known which control storage protein gene expression in maize, including Opaque2 and Opaque2 heterodimerizing protein (OHP1), which are components of a heterodimeric transcriptional activator of the zein seed storage proteins.[65] This type of regulation is not yet understood in any oilseed, but is an important biotechnological target for increasing the ratio of oil to protein in oilseeds. A putative castor homologue for OHP1 (**pCRS359**) was identified and is being further characterised by Dr. James Zhang of our laboratory.

A number of clones were identified which may have a role in protein processing and transport. These include protein disulphide isomerase (above) which

is an endoplasmic reticulum lumen protein facilitating the correct folding of proteins in the secretory pathway, including the seed storage proteins and ricin which are transported to the vacuole. Similarly, heat shock proteins (**pCRS264**, **pCRS348**, **pCRS382**, **pCRS397**, **pCRS598**, **pCRS657**, **pCRS759**, **pCRS820**, **pCRS833**; see also Table 7) are molecular chaperones involved in protein folding during normal cell metabolism and may be abundantly expressed in the developing castor seed in association with active storage protein synthesis. Storage protein transport via the golgi aparatus presumably requires quite active golgi function, and a protein involved in endoplasmic reticulum to golgi vesicle traffic (Sec12p) was homologous to one of the castor clones (**pCRS275**). Lastly, a clone was identified (**pCRS463**, see Table 8) which has homology to bacterial protein processing and/or export proteins. It is speculated that this castor clone might have a role in vacuolar protein import.

Growth of the developing castor seed tissue and storage protein deposition in the vacuoles of the growing cell requires enlargement of the vacuoles themselves. This suggests a reason for the abundance of tonoplast proteins among the clones sequenced, namely tonoplast intrinsic protein (**pCRS279**, **pCRS360**, **pCRS663**, **pCRS749**, **pCRS765**, **pCRS831**) and vacuolar ATPase (**pCRS512**, and **pCRS545**, **pCRS785**-see Table 9).

Ricinine Biosynthesis

Ricinine (*N*-methyl-4-methoxy-3-cyano-2-pyridone) accumulates in various parts of the castor plant,[66] and is derived from quinolinic acid.[67,68] Of possible

relevance to the formation of the methoxy and keto substituents are the identification of a putative O-methyltransferase (**pCRS696**) and a putative dioxygenase (**pCRS738**), respectively.

Polyamine Biosynthesis

A number of enzymes involved in polyamine biosynthesis[69] were identified. These are *S*-adenosylmethionine synthetase (**pCRS554**), a stolon tip-induced protein from potato tentatively identified[70] as *S*-adenosylmethionine decarboxylase (**pCRS555**), spermidine synthase (**pCRS763**, see Table 9), and enzymes of arginine biosynthesis (**pCRS519** and **pCRS772**, see Table 8). Polyamines are essential components of all cells, though their functions *in vivo* remain unclear[69]. Increased synthesis of polyamines appears to correlate with active growth,[69,71,72] and the putative *S*-adenosylmethionine decarboxylase of potato was identified by differential screening for clones induced in the stolon tip at the onset of tuberisation, which undergoes rapid growth.[70] Polyamines such as putrescine are precursors of some alkaloids, for example nicotine is synthesised from putrescine and nicotinic acid.[73] The castor seed accumulates the alkaloid ricinine, but this is a pyridine alkaloid, the biosynthesis of which probably does not involve polyamines.[66,67]

Organ-Specific Expression

Several clones appear to be homologous to known embryo- or seed-specific proteins. These matches are to trithorax (**pCRS282**, see Table 9), MO25 (**pCRS347**,

see Table 9), embryonic gene BP8 of birch (**pCRS371**), and embryonic abundant protein of faba bean (**pCRS565**). However, in the latter case, there is also a weaker, but apparently still homologous (since residues conserved are a subset of those conserved with the bean protein) match with "shoot-specific protein" of pea.

Contradictions of supposed specificity are observed in two other cases. "Pollen-specific" proteins of maize, rice, tomato and olive are homologous to **pCRS589**. A clue to this apparent contradiction is that the olive protein is also described as the major pollen allergen, while the castor seed is known to contain allergenic proteins. In the second example of an apparently contradictory match, two clones (**pCRS406**, **pCRS430**) had homology to "lupin-specific" clone pplzo2. Not surprisingly, the evidence for specificity to the lupin plant was very feeble.[74]

Defense-Related Proteins

Seeds typically contain various proteins presumed to have a role in defense against microbial or insect attack, and the castor seed appears to be no exception, its other defenses (such as ricin and ricinine) notwithstanding. A number of protease inhibitor clones (**pCRS413**, **pCRS648**, **pCRS653**, **pCRS754**) and thaumatin-like clones (**pCRS580**, **pCRS711**) were identified.

Dessication-Related Proteins

Two clones (**pCRS521**, **pCRS557**) identify sequences characteristic of dessication-induced proteins in other plants, typical of the later stages of seed

development.

Database Accession Numbers of Sequence Data

DNA sequences generated in this study have been deposited in the NCBI database, dbEST (database for Expressed Sequence Tags), as identification numbers 39704-40169, and in GenBank, as accession numbers T14820-T15266. Correspondence between the clone numbers used here, dbEST identification numbers, and GenBank accession numbers is depicted in Table 12.

Table 12. Database (dbEST, GenBank) accession numbers of sequences obtained in this study, corresponding to clone numbers used in the text.

dbEST	Clone	GenBank	dbEST	Clone	GenBank
39704	crs262	T14820	39741	crs304	T14857
39705	crs263	T14821	39742	crs305	T14858
39706	crs264	T14822	39743	crs309	T14859
39707	crs265	T14823	39744	crs310	T14860
39708	crs266	T14824	39745	crs311	T14861
39709	crs267	T14825	39746	crs312	T14862
39710	crs268	T14826	39747	crs313	T14863
39711	crs269	T14827	39748	crs314	T14864
39712	crs270	T14828	39749	crs315	T14865
39713	crs273	T14829	39750	crs316	T14866
39714	crs274	T14830	39751	crs317	T14867
39715	crs275	T14831	39752	crs319	T14868
39716	crs279	T14832	39753	crs320	T14869
39717	crs280	T14833	39754	crs321	T14870
39718	crs281	T14834	39755	crs322	T14871
39719	crs282	T14835	39756	crs323	T14872
39720	crs283	T14836	39757	crs324	T14873
39721	crs284	T14837	39758	crs325	T14874
39722	crs285	T14838	39759	crs328	T14875
39723	crs286	T14839	39760	crs330	T14876
39724	crs287	T14840	39761	crs331	T14877
39725	crs288	T14841	39762	crs332	T14878
39726	crs289	T14842	39763	crs333	T14879
39727	crs290	T14843	39764	crs334	T14880
39728	crs291	T14844	39765	crs335	T14881
39729	crs292	T14845	39766	crs336	T14882
39730	crs293	T14846	39767	crs337	T14883
39731	crs294	T14847	39768	crs339	T14884
39732	crs295	T14848	39769	crs340	T14801
39733	crs296	T14849	39770	crs341	T14802
39734	crs297	T14850	39771	crs342	T14803
39735	crs298	T14851	39772	crs343	T14804
39736	crs299	T14852	39773	crs345	T14805
39737	crs300	T14853	39774	crs346	T14806
39738	crs301	T14854	39775	crs347	T14807
39739	crs302	T14855	39776	crs348	T14808
39740	crs303	T14856	39777	crs349	T14809

Table 12 (cont'd). 162

dbEST	Clone	GenBank	dbEST	Clone	GenBank
39778	crs350	T14810	39819	crs398	T14916
39779	crs351	T14811	39820	crs399	T14917
39780	crs352	T14812	39821	crs400	T14918
39781	crs353	T14813	39822	crs401	T14919
39782	crs354	T14814	39823	crs403	T14920
39783	crs355	T14815	39824	crs404	T14921
39784	crs356	T14816	39825	crs405	T14922
39785	crs358	T14817	39826	crs406	T14923
39786	crs359	T14818	39827	crs407	T14924
39787	crs360	T14819	39828	crs408	T14925
39788	crs361	T14885	39829	crs409	T14926
39789	crs362	T14886	39830	crs410	T14927
39790	crs363	T14887	39831	crs411	T14928
39791	crs364	T14888	39832	crs412	T14929
39792	crs366	T14889	39833	crs413	T14930
39793	crs368	T14890	39834	crs414	T14931
39794	crs369	T14891	39835	crs415	T14932
39795	crs370	T14892	39836	crs416	T14933
39796	crs371	T14893	39837	crs417	T14934
39797	crs372	T14894	39838	crs419	T14935
39798	crs373	T14895	39839	crs420	T14936
39799	crs374	T14896	39840	crs421	T14937
39800	crs375	T14897	39841	crs424	T14938
39801	crs376	T14898	39842	crs425	T14939
39802	crs377	T14899	39843	crs426	T14940
39803	crs378	T14900	39844	crs427	T14941
39804	crs380	T14901	39845	crs428	T14942
39805	crs382	T14902	39846	crs430	T14943
39806	crs384	T14903	39847	crs431	T14944
39807	crs385	T14904	39848	crs432	T14945
39808	crs386	T14905	39849	crs433	T14946
39809	crs387	T14906	39850	crs434	T14947
39810	crs388	T14907	39851	crs435	T14948
39811	crs389	T14908	39852	crs436	T14949
39812	crs391	T14909	39853	crs437	T14950
39813	crs392	T14910	39854	crs438	T14951
39814	crs393	T14911	39855	crs439	T14952
39815	crs394	T14912	39856	crs440	T14953
39816	crs395	T14913	39857	crs441	T14954
39817	crs396	T14914	39858	crs442	T14955
39818	crs397	T14915	39859	crs443	T14956

Table 12 (cont'd).

dbEST	Clone	GenBank	dbEST	Clone	GenBank
39860	crs444	T14957	39901	crs512	T14998
39861	crs445	T14958	39902	crs513	T14999
39862	crs446	T14959	39903	crs514	T15000
39863	crs447	T14960	39904	crs515	T15001
39864	crs448	T14961	39905	crs516	T15002
39865	crs449	T14962	39906	crs517	T15003
39866	crs450	T14963	39907	crs518	T15004
39867	crs451	T14964	39908	crs519	T15005
39868	crs452	T14965	39909	crs520	T15006
39869	crs453	T14966	39910	crs521	T15007
39870	crs454	T14967	39911	crs522	T15008
39871	crs455	T14968	39912	crs523	T15009
39872	crs456	T14969	39913	crs524	T15010
39873	crs457	T14970	39914	crs539	T15011
39874	crs458	T14971	39915	crs540	T15012
39875	crs459	T14972	39916	crs541	T15013
39876	crs460	T14973	39917	crs542	T15014
39877	crs461	T14974	39918	crs543	T15015
39878	crs463	T14975	39919	crs544	T15016
39879	crs464	T14976	39920	crs545	T15017
39880	crs466	T14977	39921	crs546	T15018
39881	crs467	T14978	39922	crs547	T15019
39882	crs469	T14979	39923	crs548	T15020
39883	crs470	T14980	39924	crs549	T15021
39884	crs493	T14981	39925	crs550	T15022
39885	crs494	T14982	39926	crs551	T15023
39886	crs495	T14983	39927	crs552	T15024
39887	crs496	T14984	39928	crs553	T15025
39888	crs497	T14985	39929	crs554	T15026
39889	crs498	T14986	39930	crs555	T15027
39890	crs499	T14987	39931	crs556	T15028
39891	crs500	T14988	39932	crs557	T15029
39892	crs501	T14989	39933	crs558	T15030
39893	crs502	T14990	39934	crs559	T15031
39894	crs503	T14991	39935	crs560	T15032
39895	crs504	T14992	39936	crs562	T15033
39896	crs505	T14993	39937	crs563	T15034
39897	crs506	T14994	39938	crs564	T15035
39898	crs507	T14995	39939	crs565	T15036
39899	crs509	T14996	39940	crs566	T15037
39900	crs511	T14997	39941	crs567	T15038

Table 12 (cont'd). 164

dbEST	Clone	GenBank	dbEST	Clone	GenBank
39942	crs568	T15039	39983	crs720	T15080
39943	crs569	T15040	39984	crs721	T15081
39944	crs570	T15041	39985	crs722	T15082
39945	crs571	T15042	39986	crs724	T15083
39946	crs578	T15043	39987	crs725	T15084
39947	crs579	T15044	39988	crs727	T15085
39948	crs580	T15045	39989	crs728	T15086
39949	crs581	T15046	39990	crs729	T15087
39950	crs583	T15047	39991	crs730	T15088
39951	crs584	T15048	39992	crs731	T15089
39952	crs585	T15049	39993	crs733	T15090
39953	crs586	T15050	39994	crs734	T15091
39954	crs587	T15051	39995	crs735	T15092
39955	crs588	T15052	39996	crs736	T15093
39956	crs589	T15053	39997	crs737	T15094
39957	crs590	T15054	39998	crs738	T15095
39958	crs592	T15055	39999	crs739	T15096
39959	crs593	T15056	40000	crs741	T15097
39960	crs594	T15057	40001	crs742	T15098
39961	crs595	T15058	40002	crs743	T15099
39962	crs596	T15059	40003	crs744	T15100
39963	crs598	T15060	40004	crs745	T15101
39964	crs599	T15061	40005	crs746	T15102
39965	crs700	T15062	40006	crs747	T15103
39966	crs701	T15063	40007	crs748	T15104
39967	crs702	T15064	40008	crs749	T15105
39968	crs703	T15065	40009	crs750	T15106
39969	crs704	T15066	40010	crs751	T15107
39970	crs705	T15067	40011	crs752	T15108
39971	crs706	T15068	40012	crs753	T15109
39972	crs707	T15069	40013	crs756	T15110
39973	crs708	T15070	40014	crs757	T15111
39974	crs709	T15071	40015	crs758	T15112
39975	crs710	T15072	40016	crs759	T15113
39976	crs711	T15073	40017	crs763	T15114
39977	crs712	T15074	40018	crs764	T15115
39978	crs713	T15075	40019	crs765	T15116
39979	crs714	T15076	40020	crs766	T15117
39980	crs715	T15077	40021	crs767	T15118
39981	crs717	T15078	40022	crs768	T15119
39982	crs718	T15079	40023	crs770	T15120

Table 12 (cont'd).

dbEST	Clone	GenBank	dbEST	Clone	GenBank
40024	crs771	T15121	40065	crs645	T15162
40025	crs772	T15122	40066	crs647	T15163
40026	crs773	T15123	40067	crs648	T15164
40027	crs774	T15124	40068	crs649	T15165
40028	crs775	T15125	40069	crs650	T15166
40029	crs776	T15126	40070	crs651	T15167
40030	crs778	T15127	40071	crs652	T15168
40031	crs779	T15128	40072	crs653	T15169
40032	crs780	T15129	40073	crs654	T15170
40033	crs781	T15130	40074	crs655	T15171
40034	crs782	T15131	40075	crs656	T15172
40035	crs783	T15132	40076	crs657	T15173
40036	crs784	T15133	40077	crs658	T15174
40037	crs785	T15134	40078	crs659	T15175
40038	crs786	T15135	40079	crs660	T15176
40039	crs787	T15136	40080	crs661	T15177
40040	crs788	T15137	40081	crs662	T15178
40041	crs789	T15138	40082	crs663	T15179
40042	crs790	T15139	40083	crs666	T15180
40043	crs791	T15140	40084	crs667	T15181
40044	crs792	T15141	40085	crs668	T15182
40045	crs793	T15142	40086	crs669	T15183
40046	crs794	T15143	40087	crs670	T15184
40047	crs796	T15144	40088	crs671	T15185
40048	crs797	T15145	40089	crs672	T15186
40049	crs798	T15146	40090	crs673	T15187
40050	crs601	T15147	40091	crs674	T15188
40051	crs602	T15148	40092	crs675	T15189
40052	crs628	T15149	40093	crs676	T15190
40053	crs629	T15150	40094	crs677	T15191
40054	crs630	T15151	40095	crs678	T15192
40055	crs631	T15152	40096	crs679	T15193
40056	crs633	T15153	40097	crs680	T15194
40057	crs634	T15154	40098	crs681	T15195
40058	crs635	T15155	40099	crs682	T15196
40059	crs636	T15156	40100	crs683	T15197
40060	crs637	T15157	40101	crs684	T15198
40061	crs640	T15158	40102	crs685	T15199
40062	crs641	T15159	40103	crs686	T15200
40063	crs642	T15160	40104	crs689	T15201
40064	crs643	T15161	40105	crs690	T15202

Table 12 (cont'd).

dbEST	Clone	GenBank
40106	crs691	T15203
40107	crs692	T15204
40108	crs693	T15205
40109	crs694	T15206
40110	crs696	T15207
40111	crs697	T15208
40112	crs698	T15209
40113	crs699	T15210
40114	crs803	T15211
40115	crs804	T15212
40116	crs805	T15213
40117	crs806	T15214
40118	crs807	T15215
40119	crs808	T15216
40120	crs810	T15217
40121	crs811	T15218
40122	crs812	T15219
40123	crs813	T15220
40124	crs814	T15221
40125	crs815	T15222
40126	crs816	T15223
40127	crs817	T15224
40128	crs818	T15225
40129	crs820	T15226
40130	crs821	T15227
40131	crs822	T15228
40132	crs823	T15229
40133	crs824	T15230
40134	crs825	T15231
40135	crs826	T15232
40136	crs827	T15233
40137	crs828	T15234
40138	crs829	T15235
40139	crs831	T15236
40140	crs832	T15237
40141	crs833	T15238
40142	crs834	T15239
40143	crs835	T15240
40144	crs836	T15241
40145	crs837	T15242
40146	crs838	T15243
40147	crs839	T15244
40148	crs840	T15245
40149	crs841	T15246
40150	crs842	T15247
40151	crs843	T15248
40152	crs844	T15249
40153	crs848	T15250
40154	crs849	T15251
40155	crs852	T15252
40156	crs853	T15253
40157	crs854	T15254
40158	crs855	T15255
40159	crs856	T15256
40160	crs857	T15257
40161	crs858	T15258
40162	crs859	T15259
40163	crs860	T15260
40164	crs861	T15261
40165	crs862	T15262
40166	crs863	T15263
40167	crs864	T15264
40168	crs865	T15265
40169	crs754	T15266

CONCLUSIONS

The objective of this experiment was to identify a putative oleate-12-hydroxylase cDNA, and this was achieved. The two clones pCRS677 and pCRS834 have sequence similarity with plant membrane-bound desaturase genes, but do not encode a previously identified desaturase. These clones are examined in detail in the subsequent chapter.

The secondary objective of this experiment was to isolate other ovel plant genes, and this also was achieved. A large proportion of the clones sequenced could be identified by comparison to sequence databases, which may reflect the relatively high levels of expression in the developing castor seed of relatively well characterised enzymes, such as those of glycolysis. Among the sequences identified were a coniderable number which appear to be previously unknown or uncharacterised in higher plants, and these are expected to provide new oportunities in plant biology.

REFERENCES

1. Coppens, N., Biosynthesis of fatty acids by seeds of *Ricinus communis, Nature,* 177, 279, 1956.

2. Canvin, D.T., Formation of oil in the seed of Ricinus communis L., *Can. J. Biochem. Physiol.*, 41, 1879-1885, 1963.

3. Yamada, M., and Stumpf, P.K., Enzymic synthesis of ricinoleic acid by extracts of developing Ricinus communis L. seeds, *Biochem. Biophys. Res. Commun.*, 14, 165-171, 1964.

4. James, A.T. Hadaway, H.C., and Webb, J.P.W., The biosynthesis of ricinoleic

acid, *Biochem. J.*, 95, 448-452, 1965.

5. Hajela, R.K., Horvath, D.P., Gilmour, S.J., Thomashow, M.F., Molecular cloning and expression of *cor* (cold-regulated) genes in *Arabidopsis thaliana*, *Plant Physiol.*, 93, 1246-1252, 1990.

6. Shanklin, J. and Somerville, C., Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs, *Proc. Natl. Acad. Sci. USA*, 88, 2510, 1991.

7. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

8. Kamalay, J.C., and Goldberg, R.B., Regulation of structural gene expression in tobacco, *Cell*, 19, 935-946, 1980.

9. Kamalay, J.C., and Goldberg, R.B., Organ-specific nuclear RNAs in tobacco, *P.N.A.S. USA*, 81, 2801-2805, 1984.

10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., Basic local alignment search tool, *J. Mol. Biol.*, 215, 403-410, 1990.

11. Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C., and Venter, J.C., Sequence identification of 2375 human brain genes, *Nature*, 355, 632-634, 1992.

12. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C., Complementary DNA sequencing : expressed sequence tags and human genome project, *Science*, 252, 1651-1656, 1991.

13. Keith, C.S., Hoang, D.O., Barrett, B.M., Feigelman, B., Nelson, M.C., Thai, H., and Bayersdorfer, C., Partial sequence analysis of 130 randomly selected maize cDNA clones, *Plant Physiol.*, 101, 329-332, 1993.

14. Uchimiya, H., Kidou, S., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nishi, R., Hashimoto, H., Matsubayashi, Y., Kidou, N., Umeda, M., and Kato, A., Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.), *Plant J.*, 2, 1005-1009, 1992.

15. Hoog, C., Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy, *Nucleic Acids Research*, 19, 6123-6127, 1991.

16. Parsot, C., Evolution of biosynthetic pathways: a common ancestor for threonine synthase, threonine dehydratase and D-serine dehydratase, *EMBO J.*, 5, 3013-3019, 1986.

17. Anselme, J., and Hartlein, M., Tyr-426 of the *Escherichia coli* asparaginyl-tRNA synthetase, an amino acid in a conserved C-terminal motif, is involved in ATP binding, *FEBS Lett.*, 280, 163-166, 1991.

18. Anselme, J., and Hartlein, M., Asparaginyl-tRNA synthetase from *Escherichia coli* has significant sequence homologies with yeast aspartyl-tRNA synthetase, *Gene*, 84, 481-485, 1989.

19. Youvan, D.C., Bylina, E.J., Alberti, M., Begusch, H., and Hearst, J.E., Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from R. capsulata, *Cell*, 37, 949-957, 1984.

20. Bartsch, K., Johnn-Marteville, A. von, and Schulz, A., Molecular analysis of two genes of the *Escherichia coli gab* cluster: nucleotide sequence of the glutamate:succinic semialdehyde transaminase gene (*gabT*) and characterisation of the succinic semialdehyde gene (*gabD*), *J. Bacteriol.*, 172, 7035-7042, 1990.

21. Matsubara, Y., Indo, Y., Naito, E., Ozasa, H., Glassberg, R., Vockley, J., Ikeda, Y., Kraus, J., and Tanaka, K., Molecular cloning and nucleotide sequence of cDNAs encoding the precursors of rat long chain acyl-coenzyme A, short-chain acyl-coenzyme A, and isovaleryl-coenzyme A dehydrogenases; sequence homology of four enzymes of the acyl-CoA dehydrogenase family, *J. Biol. Chem.*, 264, 16321-16331, 1989.

22. Komine, Y., and Inokuchi, H., Precise mapping of the *rnpB* gene encoding the RNA component of RNaseP in *Escherichia coli* K-12, *J. Bacteriol.*, 173, 1813-1816, 1991.

23. Ohlrogge, J.B., Jaworski, J.G., and Post-Beittenmiller, D., *De novo* fatty acid biosynthesis, in *Lipid Metabolism in Plants*, Moore, T.S., Ed., CRC Press, Boca Raton, 1993, chap. 1.

24. Sparace, S.A., and Kleppinger-Sparace, K.F., Metabolism in nonphotosynthetic, nonoilseed tissues, in *Lipid Metabolism in Plants*, Moore, T.S., Ed., CRC Press, Boca Raton, 1993, chap. 17.

25. Petrovics, G., Putnoky, P., Kim, J., Thorp, T.A., Noel,. K.D., Carlson, R.W., and Kondorosi, A., The presence of a novel type of surface

polysaccharide in *Rhizobium meliloti* requires a new fatty acid synthase-like gene cluster involved in symbiotic nodule development, *Molecular Microbiology,* 8, 1083-1094, 1993.

26. Mathur, M., and Kolattukudy, P.E., Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. *bovis Bacillus Calmette-Guerin, J. Biol. Chem.*, 267, 19388-19395, 1992.

27. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L., Modular organization of genes required for complex polyketide biosynthesis, *Science,* 252, 675-679, 1991.

28. Coleman, J., and Raetz, C.R.H., First committed step of lipid A biosynthesis in *Escherichia coli*: sequence of the *lpxA* gene, *J. Bacteriol.*, 170, 1268-1274, 1988.

29. Crowell, D.N., Reznikoff, W.S., and Raetz, C.R.H., Nucleotide sequence of the *Escherichia coli* gene for lipid A disaccharide synthase, *J. Bacteriol.*, 169, 5727-5734, 1987.

30. Tomasiewicz, H.G., and McHenry, C.S., Sequence analysis of the *Escherichia coli dnaE* gene, *J. Bacteriol.*, 169, 5735-5744, 1987.

31. Raetz, C.R.H., Biochemistry of endotoxins, *Annu. Rev. Biochem.*, 59, 129-170, 1990.

32. Holst, O., Muller-Loennies, S., Lindner, B., and Brade, H., Chemical structure of the lipid A of *Escherichia coli* J-5, *Eur. J. Biochem.*, 214, 695-701, 1993.

33. Helander, I.M., Lindner, B., Seydel, U., and Vaara, M., Defective biosynthesis of the lipid A component of temperature-sensitive *firA* (*omsA*) mutant of *Escherichia coli, Eur. J. Biochem.*, 212, 363-369, 1993.

34. Anderson, M.S., and Raetz, C.R.H., UDP-*N*-Acetylglucosamine 3-*O*-Acyltransferase from *Escherichia coli, Meth. Enzymol.*, 209, 449-454, 1992.

35. Raetz, C.R.H., Lipid A disaccharide synthase from *Escherichia coli, Meth. Enzymol.*, 209, 455-466, 1992.

36. Brozek, K.A., and Raetz, C.R.H., Biosynthesis of lipid A in *Escherichia coli*: acyl carrier protein-dependent incorporation of laurate and myristate, *J. Biol. Chem.*, 265, 15410-15417, 1990.

37. Brozek, K.A., and Raetz, C.R.H., 3-Deoxy-D-*manno*-octulosonate transferase and late acyltransferases of lipopolysaccharide biosynthesis, *Meth. Enzymol.*, 209, 476-485, 1992.

38. Cronan, J.E., Jr., Li, W.-B., Coleman, R., Narasimhan, M., Mendoza, D. de, and Schwab, J.M., Derived amino acid sequence and identification of active site residues of *Escherichia coli* β-hydroxydecanoyl thioester dehydrase, *J. Biol. Chem.*, 263, 4641-4646, 1988.

39. Shimakata, T., and Stumpf, P.K., Purification and characterizations of β-ketoacyl-[acyl carrier protein] reductase, β-hydroxyacyl-[acyl carrier protein] dehydrase, and enoyl-[acyl carrier potein] reductase from *Spinacia oleracea* leaves, *Arch. Biochem. Biophys.*, 218, 77-91, 1982.

40. Foster, S.J., Purification and characterization of an "actomyosin" complex from Escherichia coli W3110, *FEMS Microbiol. Lett.*, 110, 295-298, 1993.

41. Sasaki, Y., Hakamada, K., Suama, Y., Nagano, Y., Furusawa, I., and Matsuno, R., Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant, *J. Biol. Chem.*, 268, 25118-25123, 1993.

42. Tsinoremas, N.F., Castets, A.M., Harrison, M.A., Allen, J.F., and Tandeau de Marsac, N., Photosythetic electron transport controls nitrogen assimilation in cyanobacteria by means of posttranslational modification of the *glnB* gene product, *Proc. Natl. Acad. Sci. USA*, 88, 4565-4569, 1991.

43. Edwards, J.W., and Coruzzi, G.M., Photorespiration and light act in concert to regulate the expression of the nuclear gene for chloroplast glutamine synthetase, *Plant Cell*, 1, 241-248, 1989.

44. Garrard, L.J., and Goodman, J.M., Two genes encode the major membrane-associated protein of methanol-induced peroxisomes from *Candida boidinii*, *J. Biol. Chem.*, 264, 13929-13937, 1989.

45. Gould, S.J., Keller, G.-A., Schneider, M., Howell, S.H., Garrard, L.J., Goodman, J.M., Distel, B., Tabak, H., and Subramani, S., Peroxisomal protrin import is conserved between yeast, plants, insects and mammals, *EMBO J.*, 9, 85-90, 1990.

46. Gould S.J., Keller, G.-A., Hosken, N., Wilkinson, J., and Subramani, S., A conserved tripeptide sorts proteins to peroxisomes, *J. Cell Biol.*, 108, 1657-1664, 1989.

47. Olsen, L.J., Ettinger, W.F., Damsz, B., Matsudaira, K., Webb, M.A.,

Harada, J.J., Targetting of glyoxisomal proteins to peroxisomes in leaves and roots of a higher plant, *Plant Cell*, 5, 941-952, 1993.

48. Mazo, A.M., Huang, D., Mozer, B.A., and Dawid, I.B., The trithorax gene, a trans-acting regulator of the bithorax complex in *Drosophila*, encodes a protein with zincbinding domains, *Proc. Natl. Acad. Sci. USA*, 87, 2112-2116, 1990.

49. Tkachuk, D.C., Kohler, S., and Cleary, M.L., Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias, *Cell*, 71, 691-700.

50. Miyamoto, H., Matsushiro, M., and Nozaki, M., Molecular cloning of a novel mRNA sequence expressed in cleavage stage mouse embryos, *Molecular Reproduction and Development*, 34, 1-7, 1993.

51. Kushnirov, V.V., Ter-Avanesyan, M.D., Didichenko, S.A., Smirnov, V.N., Chernoff, Y.O., Derkach, I.I., Novikova, O.N., Inge-Vechtomov, G.I., Neistat, M.A., and Tolstorukov, I.I., Divergence and conservation of *sup2 (sup35)* gene of yeasts *Pichia pinus* and *Saccharomyces cerevisiae*, *Yeast*, 6, 461-472, 1990.

52. Hoshino, S., Miyazawa, H., Enomoto, T., Hanaoka, F., Kikuchi, Y., Kikuchi, A., and Ui, M., A human homologue of the yeast *GST1* gene codes for a GTP-binding protein and is expressed in a proliferation-dependent manner in mammalian cells, *EMBO J.*, 8, 3807-3814, 1989.

53. Lorenz, R.T., and Parks, L.W., Cloning, sequencing, and disruption of the gene encoding sterol C-14 reductase in *Saccharomyces cerevisiae*, *DNA Cell Biol.*, 11, 685-692, 1992.

54. Shimanuki, M., Goebl, M., Yanagida, M., and Toda, T., Fission yeast *sts1*$^+$ gene encodes a protein similar to the chicken lamin B receptor and is impicated in pleiotropic drug-sensitivity, divalent cation-sensitivity, and osmoregulation, *Mol. Biol. Cell*, 3, 263-273, 1992.

55. Sibille, C., Chomez, P., Wildmann, C., Van Pel, A., De Plaen, E., Maryanski, J.L., Bergeyck, V de, and Boon, T., Structure of the gene of tum$^-$ transplantation antigen P198: a point mutation generates a new antigenic peptide, *J. Exp. Med.*, 172, 35-45, 1990.

56. Werner, A., Moore, M.L., Mantei, N., Biber, J., Semenza, G., and Murer, H., Cloning and expression of cDNA for a Na/Pi cotransport system of kidney cortex, *Proc. Natl. Acad. Sci.*, 88, 9608-9612, 1991.

57. Poirier, Y, Thoma, S., Somerville, C., and Schieffelbein, J., A mutant of *Arabidopsis* deficient in xylem loading of phosphate, *Plant Physiol.*, 97, 1087-1093, 1991.

58. Rokeach, L.A., Haselby, J.A., and Hoch, S.O., Molecular cloning of a cDNA encoding the human Sm-D autoantigen, *Proc. Natl. Acad. Sci. USA*, 85, 4832-4836, 1988.

59. Raikhel, N., Nuclear targetting in plants, *Plant Physiol.*, 100, 1627-1632, 1992.

60. Lord, J.M., Lamb, I.F., and Roberts, L.N., Ricin: structure, biological activity and synthesis, *Oxford Surveys Plant Molec. Cel Biol.*, 1, 85-101, 1984.

61. Sullivan, T.D., Strelow, L.I., Illingworth, C.A., Phillips, R.L., and Nelson, O.E., Jr., Analysis of maize *Brittle-1* alleles and a defective *suppressor-mutator*-induced mutable allele, *Plant Cell*, 3, 1337-1348, 1991.

62. Smith, R.G., Gauthier, D.A., Dennis, D.T., and Turpin, D.H., Malate- and pyruvate-dependent fatty acid synthesis in leucoplasts from developing castor endosperm, *Plant Physiol.*, 98, 1233, 1992.

63. Wetterau, J.R., Combs, K.A., Spinner, S.N., and Joiner, B.J., Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex, *J. Biol. Chem.*, 265, 9800-9807, 1990.

64. Sharp, D., Blinderman, L., Combs, K.A., Kienzle, B., Ricci, B., Wager-Smith, K., Gil, C.M., Turck, C.W., Bouma, M.-E., Rader, D.J., Aggerbeck, L.P., Gregg, R.E., Gordon, D.A., and Wetterau, J.R., Cloning and gene defects in microsomal triglyceride transfer protein associated with abetalipoproteinaemia, *Nature*, 365, 65-69, 1993.

65. Pysh, L.D., Aukerman, M.J., and Schmidt, R.J., OHP1: a maize basic domain/leucine zipper protein that interacts with Opaque2, *Plant Cell*, 5, 227-236, 1993.

66. Rizk, A.-F.M., and El-Missiry, M.M., Non-diterpenoid constituents of Euphorbiaceae and Thymelaeaceae, in *Naturally Ocurring Phorbol Esters*, Evans, F.J., Ed., CRC Press, Boca Raton, 1986, chap 5.

67. Hiles, R.A., The chemistry and biochemistry of ricinine and some related pyridine compounds, PhD Thesis, Department of Biochemistry, Michigan State University, 1970.

68. Yang, K.S., and Waller, G.R., Biosynthesis of the pyridine ring of ricinine from quinolinic acid glycerol and aspartic acid, *Phytochem.*, 4, 881-889, 1965.

69. Galston, A.W. and Kaur-Sawhney, R., Polyamines as endogenous growth regulators, in *Plant Hormones and Their Role in Plant Growth and Development,* Davies, P.J., Ed., Martinus Nijhof, Dordrecht, 1987, 280-295.

70. Taylor, M.A., Mad Arif, S.A., Kumar, A., Davies, H.V., Scobie, L.A., Pearce, S.R., and Flavell, A.J., Expression and sequence analysis of cDNAs induced during the early stages of tuberisation in different organs of the potato plant (*Solanum tuberosum* L.), *Plant Mol. Biol.*, 20, 641-651, 1992.

71. Smith, T.A., Polyamines, *Ann. Rev. Plant Physiol.*, 36, 117-143, 1985.

72. Evans, P.T., and Malmberg, R.L., Do polyamines have roles in plant development?, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 40, 235-269, 1989.

73. Hamill, J.D., Robins, R.J., Parr, A.J., Evans, D.M., Furze, J.M., and Rhodes, M.J.C., Over-expressing a yeast ornithine decarboxylase gene in transgenic roots of *Nicotiana rustica* can lead to enhanced nicotine accumulation, *Plant Mol. Biol.*, 15, 27-38, 1990.

74. Perrey, R., Warskulat, U., and Wink, M., Molecular cloning of a lupin-specific gene from a cDNA library of suspension-cultured cells of *Lupinus polyphyllus*, *Plant Mol. Biol.*, 15, 175-176, 1990.

CHAPTER 6

ANALYSIS OF PUTATIVE OLEATE 12-HYDROXYLASE CLONES

ABSTRACT

Two putative oleate-12-hydroxylase clones, pCRS677 and pCRS834, identified in chapter 5, were compared and found to contain identical sequences. A full length clone, pFL2, was isolated using pCRS677 as a probe, and sequenced. An 1161 bp open reading frame encoding a 44407 Da protein had limited sequence similarity (37% at the amino acid level) to plant membrane-bound desaturases. Northern blot analysis revealed that pFL2 corresponds to a strongly-expressed seed-specific transcript in castor. A Southern blot showed that this transcript originates from a single-copy gene. The clone pFL2 was engineered for expression in transgenic yeast and plants. Three plant transformation systems were employed: tobacco leaf explants, tobacco cultured cells, and carrot root disks were each transformed by co-cultivation with *Agrobacterium tumefaciens*. No ricinoleic acid could be detected in any of the transgenic tissues, nor were there any consistent changes in the relative proportions of the usual fatty acids. It is concluded that the sequence and pattern of expression of pFL2 are consistent with the possibility that it encodes oleate-12-hydroxylase, but this was neither supported nor refuted by attempts at expression in transgenic yeast and plants.

INTRODUCTION

In the preceding chapter, an experiment is described in which seed-specific clones expressed at moderate levels of abundance were selected by differential screening, and a large number of these clones were partially sequenced. The principle objective of this experiment was to isolate a gene encoding oleate-12-hydroxylase. Since it was hypothesized that this gene is homologous to those of the microsomal fatty acid desaturases, it was hoped that the oleate-12-hydroxylase gene could be identified on this basis. Two clones, pCRS677 and pCRS834, gave such a match to the *Brassica napus fad3* gene, encoding the microsomal ω-3 desaturase, the only known higher plant membrane-bound desaturase which was in the databases. These two clones were therefore considered putative oleate-12-hydroxylase clones and their analysis is the subject of this chapter.

The partial clone pCRS677 was used to isolate full-length cDNA clones, one of which was completely sequenced. Since ricinoleic acid is only produced in the seed,[1-4] it was expected that a hydroxylase clone should only be expressed in seed tissues, and this was examined by northern analysis. In order to test whether the isolated clone encodes a protein with oleate-12-hydroxylase activity, the clone was expressed in both yeast and plants.

MATERIALS AND METHODS

Screening of pYES2.0 cDNA Library

Three plates of each of the four pools of the pYES2.0 cDNA library (chapter 3) were screened by the same method described above (chapter 4). The pCRS677 insert was excised with *Bam*HI and *Apa*I, gel-purified, ^{32}P-labelled by random priming[5] and purified of unincorporated nucleotides by ethanol precipitation in the presence of ammonium acetate.[5] This probe was hybridised to the filters overnight at 65°C, and unhybridised probe was removed by washing at room temperature (using the method and solutions described in chapter4).

DNA Sequencing

Terminal sequencing of various clones in pYES2.0 was as described above (chapter 4) using the T7 fluorescent primer, or the primer F1 (5' AGC GTG ACA TAA CTA ATT 3') with fluorescent terminators. Sequencing of the entire pFL2 gene employed, in addition to F1 and T7, the following oligonucleotide primers, in combination with fluorescent terminators:

HF2: 5' GCT CTT TTG TGC GCT CAT TC 3'

HF3: 5' GTC CAT TCT GCA CTT CTG GT 3'

HF4: 5' ACG ATC GCT TTG CTT GCC AT 3'

HF5: 5' GGA GCA ATG GTG ACT GTC GA 3'

HF6: 5' CAA GGC GTT TTC TGG TAC CG 3'

HR1: 5' CGG TAC CAG AAA ACG CCT TG 3'

HR2: 5' TCG ACA GTC ACC ATT GCT CC 3'

HR3: 5' ATG GCA AGC AAA GCG ATC GT 3'

HR4-2: 5' GAG AAT GCA GCC TTG GAA GA 3'

HR5: 5' GAA TGA GCG CAC AAA AGA GC 3'

HR6: 5' CTT CAA GCG GGA GTT CAA CC 3'.

Sequence data was analysed using the programs DNASIS and PROSIS.

Northern Blot Analysis

Poly(A)$^+$ RNA prepared (chapter 3) from leaves and developing seeds was electrophoresed through an agarose gel containing formaldehyde.[5] An equal quantity (3 μg) of RNA was loaded in both lanes, and RNA standards (0.16-1.77 kb ladder, Gibco-BRL) were loaded in a third lane. Following electrophoresis, RNA was transferred from the gel to a nylon membrane (Hybond N, Amersham) and fixed to the filter by exposure to UV light for 2 min (based on an empirical calibration). A ^{32}P-labelled probe was prepared from insert DNA of clone pCRS677 as above, and hybridised to the membrane overnight at 65°C, after it had been prehybridised for ~1 h. The hybridization solution contained 4 X SET (0.6 M NaCl, 0.12 M Tris-HCl pH 7.4, 8 mM EDTA), 0.1% sodium pyrophosphate, 0.2% SDS, 0.1% heparin, and

5% dextran sulphate. The blot was washed three times in 2 X SSC, 0.1% SDS at room temperature, then exposed to X-ray film, and to a phosphor-imaging screen (Molecular Dynamics). A probe was subsequently made from the *Colletotrichum graminicola* β-tubulin gene *TUB2*[6] and hybridised to the same blot under the same conditions, except that the hybridization temperature was reduced to 58°C, and exposed to X-ray film.

Southern Blot Analysis

Genomic *Arabidopsis* DNA (1 μg, courtesy of Dr. Yves Poirier) and genomic castor DNA (2 μg, courtesy of Dr. John Shanklin) were digested with *Eco*RI, *Bam*HI, or *Hin*dIII, in reactions supplemented with spermidine (1 mM), and separated in 0.7% agarose gel. A Southern blot was prepared as described in chapter 4, and hybridised and washed in the solutions described for the northern blot, above.

Yeast Strains and Transformation

Yeast strain CGY2557 is described in chapter 3. The strain JO522 was obtained from Dr. Joe Ogas: *MATa*, GAL$^+$, *ura3-52, trp1Δ65, leu2, his3Δ1, prs1-1122, pep4-3, prc1-407.*

A method for electroporation of yeast was used that was simpler, but less efficient, than that used in chapter 3. Yeast were grown to OD_{600} = 1.0 in 100 ml

YPD medium (10 g l^{-1} yeast extract, 20 g $^{-1}$ bacto peptone, 20 g l^{-1} dextrose). Cells were pelleted (SS-34 rotor, 5000g, 5 min, 4°C) and resuspended in ~200 μl of the supernatant. These cells (40 μl per transformation) were used for electroporation with a minimal volume of DNA in cold, 2 mm-gap, cuvettes with a BioRad instrument set at 600 V, 200 Ω, 25 μF. The cells were diluted with 160 μl YPD medium and plated on selective medium (chapter 3), without sorbitol.

Gas Chromatography

Fatty acid methyl ester standards of ricinoleate (Sigma), Rapeseed Reference Mixture (Supelco), and an equal-mass mixture of 16:0, 18:0, 18:1, 18:2 and 18:3 (Sigma), were generally injected at 0.125-0.25 mg ml^{-1}.

Yeast Samples

Yeast cells (from ~1 cm^2 patches) were scaped from the plate and transferred to a glass scew-cap tube. 1.0 M methanolic HCl (1.5 ml) was added, the tube capped with a teflon-lined cap, and heated to 80°C for 1 h. Upon cooling, 1 ml hexane:isopropanol (3:2) and 0.5 ml 0.2 M Na_2SO_4 were added and the FAMEs removed in the hexane phase, which could be stored at -20°C. The samples were analysed with a Hewlet-Packard 5890 series II gas chromatograph equipped with a SPB-1 thin film fused silica capillary column (15 m long, 0.53 mm internal diameter, 0.10 μm film thickness; Supelco) with helium as carrier gas. Methyl-ricinoleate chromatographs well on this column without TMS derivatization. The injected sample was split between the column and the purge valve. The column was held at 150°C

for four minutes and then programmed to 245°C at 4°C min^{-1}.

Plant Samples

Experiment 1: The oldest part (1-2 cm) of carrot roots were used, so as to allow continued growth of the meristematic part. FAMEs were prepared as above, dried under nitrogen, redissolved in hexane (2 x 25 μl), and transferred to a gas chromatograph vial, where trimethylsilylimidazole (TriSilZ, Pierce; 0.5 μl) was added, to derivatize any hydroxyl groups in a rapid room temperature reaction. The samples were analysed on a SPB-1 fused silica capillary column (30 m long, 0.25 mm internal diameter, 0.25 μm film thickness; Supelco) using a Hewlet-Packard 5971 series mass selective detector, in place of the flame ionization detector used in other experiments. The detector scanned masses between 40 and 650. The injected sample was not split; the oven was held at an initial temperature of 150°C for 7 min, then programmed to 250°C at 5°C min^{-1}.

Experiments 2 and 3: Samples were prepared as for experiment 1, except the final sample was dissolved in 2 x 50 μl hexane, analysed by gas chromatography once before derivatization with TMS, and then injected a second time. SP2330 glass capillary columns (30 m, 0.75 mm ID, 0.20 μm film, Supelco) were used, the samples were not split, the temperature program was 150°C (6 min) to 215°C (4°C min^{-1}), and flame ionization detectors were used.

Plant Transformation

Constructs for expression of the pFL2 insert in plants using the vector pBI121[7] were prepared by two independent routes. The use of this vector, in which the only 3' cloning site is *Sac*I, was complicated by the presence of a *Sac*I site in the coding region of the pFL2 insert. In the first route, pFL2 was linearised with *Xba*I (which cuts at the 3' region flanking the insert), blunt-ended with the Klenow fragment of DNA polymerase I, then digested with *Bam*HI (which cuts at the 5' end of the insert), releasing the insert, which was gel-purified. The vector pBI121 was digested with *Sac*I and blunt-ended with T4 DNA polymerase, then cut with *Bam*HI and treated with calf intestinal phosphatase to prevent religation with the excised β-glucuronidase fragment. The pFL2 insert was ligated to this pBI121 vector and used to transform *Escherichia coli* DH5α cells to kanamycin resistance. Plasmid DNA of transformants was digested with *Xba*I and *Sac*I, and two clones (A4, B6) were chosen that had the 1.3 kb fragment indicating that the pFL2 cDNA was correctly inserted into the pBI121 vector. This was confirmed by the fact that *Sna*BI did not cut these clones (*Sna*BI cuts the β-glucuronidase gene), and *Eco*RI/*Hin*dIII released a band of appropriate size (~2.5 kb).

In the second route, clone pFL2 was digested with *Xba*I and then partially digested with *Sac*I. A band of ~1.45 kb representing the entire insert was isolated from a gel. The vector SLJ4K1 (J. Jones, unpublished: see Figure 16) was digested with *Xba*I and *Sac*I, and the vector fragment was gel-purified. The pFL2 insert was

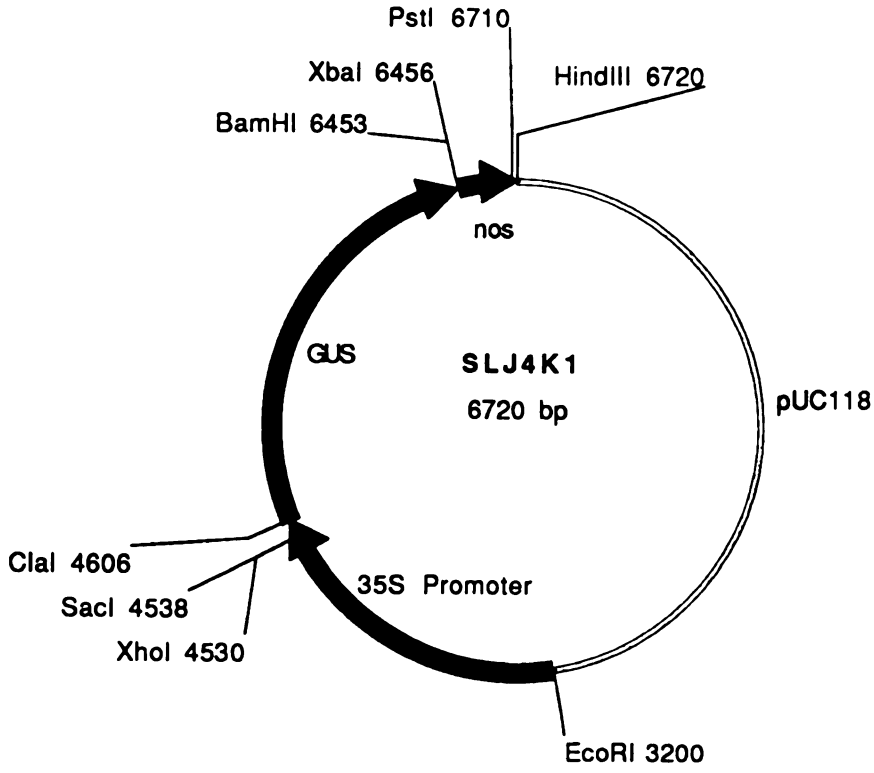


Figure 16. The vector SLJ4K1 constructed by J. Jones, and used for preparation of constructs for expression of pFL2 in plants.

ligated to this vector, transformed into DH5α, and checked for the presence of the 1.3 kb *Sac*I insert fragment. Such a clone was then digested with *Eco*RI and *Hin*dIII, and this DNA was ligated to the large *Eco*RI/*Hin*dIII fragment of pBI121, transformed into DH5α and selected for both kanamycin resistance and ampicillin sensitivity. By this procedure, the entire (35S promoter)-(pFL2 insert)-(nos terminator) fragment derived from SLJ4K1 was used to replace the (35S promoter)-(β-glucuronidase)-(nos terminator) fragment of pBI121. The clones obtained were digested with *Sac*I, and one clone (9/18 3) which gave the appropriate 1.3 kb fragment was selected.

The three clones (A4 and B6 prepared by the first route, and 9/18 3 prepared by the second), plus the unmodified vector pBI121, were transformed into *Agrobacterium tumefaciens* strains GV3101, R1000 and LBA4404 by electroporation. Cells for electroporation were prepared as follows. GV3101 and R1000 were grown in LB medium with reduced NaCl (5 g l^{-1}), and LBA4404 was grown in TY medium (5 g l^{-1} bacto-tryptone, 3 g l^{-1} yeast extract, pH 7.5). A 500 ml culture was grown to OD_{600} = 0.6, then centrifuged at 4000 rpm (GS-A rotor) for 5 min. The supernatant was aspirated immediately from the loose pellet, which was gently resuspended in 500 ml ice-cold water. The cells were centrifuged as before, resuspended in 30 ml ice-cold water, transferred to a 30 ml tube and centrifuged at 5000 rpm (SS-34 rotor) for 5 min. This was repeated three times, resuspending the cells consecutively in 30 ml ice-cold water, 30 ml ice-cold 15% dimethyl sulfoxide (DMSO), and finally in 4 ml ice-cold 15% DMSO. These cells were aliquoted, frozen in liquid nitrogen, and

stored at -80°C. Electroporations employed a BTX instrument using cold 1 mm-gap cuvettes containing 40 μl cells and a minimal volume of DNA, a voltage of 1.44 KV, and 129 Ω resistance. The electroporated cells were diluted with 1 ml SOC medium[5] and incubated at 28°C for 1-2 h before plating on medium containing kanamycin (50 mg l^{-1}.

Plasmid DNA was purified from *Agrobacterium* cells using "Magic Minipreps" (Promega) according to the instructions of the manufacturer (for *E. coli*), with the exception that lysozyme (2.7 mg ml^{-1}) was added to the resuspension solution in which cells were incubated for 10 min at room temperature, and a lysis time of 5 min was strictly observed. Plasmid yields from cells grown in TY medium were better than from cells grown in LB medium. Plasmid DNA from *Agrobacterium* cultures used to transform plant tissues was digested with restriction enzymes and compared to the plasmid DNA isolated from *E. coli*, and was in all cases identical.

Fresh carrots were obtained from the garden of Drs Susan Gibson, Deane Falcone and Joe Ogas, or from a local store, and transformed using strain R1000 according to Petit *et al*,[8] with the modifications of Yadav *et al*.[9] Peeling of the carrots (with a sterile scalpel) was essential; surface sterilization treatments with bleach alone were insufficient to prevent infection of the disks.

Nicotiana tabacum SR-1 leaf explants were transformed according to Newman *et al*,[10] except that leaves were maintained on No. 3 medium for 3 days prior to inoculation.

N. tabacum NT-1 cells were transformed according to Newman *et al*.[10]

RESULTS AND DISCUSSION

Isolation and Sequencing of cDNA Clone pFL2

Comparison of the initial sequence data of pCRS677 and pCRS834 obtained with the T3 primer, indicated that these are independent clones derived from the same gene. Clone pCRS677 has a small (14 bp) deletion in the vector/linker sequence immediately preceding the cDNA sequence, but is otherwise identical to pCRS834.

The insert of pCRS677 (~700 bp) was used as a probe to screen the pYES2.0 library (chapter 3), by colony hybridization at high stringency. In the primary screen of 47 000 colonies, 84 hybridizing colonies were obtained. The first 28 of these positive colonies were screened again, to obtain pure positive clones. All 28 of the primary positives were positive in the secondary screen, indicating an overall frequency of one positive clone per 560 clones in the library.

DNA prepared from the 28 purified clones was digested with restriction enzymes and analysed by agarose gel electrophoresis. The enzymes *Bam*HI and *Xho*I cut the vector on either side of the cloning site, and therefore should excise the inserted DNA when used together. With one exception, all clones had a single fragment smaller than ~800 bp, or an ~800 bp fragment plus one or two additional fragments. Clone 4avi did not fit this pattern. A double-digest with *Xba*I and *Hin*dIII should, similarly, excise the insert. All clones analysed yielded only one fragment, ranging in size between ~700 bp and ~2.2 kb, except clone 4avi, which had an

insert of ~4 kb. Due to minor technical difficulties, however, clones 2ci, 3cv, 4cii, 4aiii, 4ci, 4aii, 4ai, and 3ciii, were not analysed by digestion with *Xba*I and *Hin*dIII. The majority of clones had one *Hin*cII site in the insert, with the exception of clones 3cv, 4cii, 4aiii, 4ci, 3cii, 4aii, 3cvii, and 3cvi, which either lacked this site or had an additional site. Taken together, these results indicate that most of the 28 clones purified have a similar restriction pattern, with 9 possible exceptions. This is compatible with the possibility that most, if not all, represent the same gene. Those that appeared to have different restriction patterns were not analysed further. Of the majority of clones, which appeared to have similar restriction patterns but varying insert sizes, 14 were used to obtain sequence data (below). This data supports the conclusion that these 14 clones were derived from the same gene. It is concluded that this one class of clones is present in the pYES2.0 library at a frequency between 1/560 and 1/1120. This frequency is compatible with the possibility that these clones encode oleate-12-hydroxylase, which is actively expressed in the developing castor seed.

As already mentioned, 14 of the clones were used to obtain sequence data, by sequencing from one or both ends of each clone, in addition to the data previously obtained from clones pCRS677 and pCRS834. These sequence data were used to assemble a contiguous sequence of ~1.7 kb. The longest clone, 3cvii, was 113 bp longer than the next longest, 3civ-1. However, the first 305 bp of 3cvii showed no similarity to the overlapping portion of 3civ-1 or several other clones of similar length, which were, however, all identical in sequence to each other (Figure 17).

```
                    210        220        230        240        250
3cvii       234 ACTTGGTGAT GATAGTTCCG GTTATAGCAA ATCCGACCAA AAACGGCCAG   283
3civ-1       79 CACACTTGGT GACCTCAAAT CAAACACCAC ACCTTATAAC TTAGTCTTAA   128
4cvii-1      -6 .......... .......... .......... .......... CTAAAGTTAA    44
3cx-3       -56 .......... .......... .......... .......... ..........    -7
3ci         -19 .......... .......... .......... .......... ..........    31
4av           5 .......... .......... .......... .......... ..........    54
2cii          3 .......... .......... .......... .......... ..........    52
3cix        -40 .......... .......... .......... .......... ..........    10
4avii       -31 .......... .......... .......... .......... ..........    19
2ciii-4       4 .......... .......... .......... .......... ..........    53

                    260        270        280        290        300
3cvii       284 TTACGGTTGA ACTCCCGCTT GAAGAACACG GGCCATGGAT CGAACCACCT   333
3civ-1      129 GAGAGAGAGA GAGAGAGAGG AGACATTTCT CTTCTCTGAG ATAAGCACTT   178
4cvii-1      45 GAGAGAGAGA GAGAGAGAGG AGACATTTCT CTTCTCTGAG ATAAGCACTT    94
3cx-3        -6 .......... .......... .......... .......... .TCTAAAGTT    44
3ci          32 .......... .......... .......... ......CTCT AAAGGCACTT    81
4av          55 ..CTCTAAAG GAGAGAGAGG AGACATTTCT CTTCTCTGAG ATAAGCACTT   104
2cii         53 .......... .CTCTAAAGG AGACACTTCT CTTCTCTGAG ATAAGCACTT   102
3cix         11 .......... .......... .......... .......... ........CT    60
4avii        20 .......... .......... .......... ........CT CTAAAGACTT    69
2ciii-4      54 .......... ...CTCTAAA GGACATTTCT CTTCTCTGAG ATAAGCACTT   103

                    310        320        330        340        350
3cvii       334 TTTCATCTTT TCTCGAAGCC TCAGGAAAGT GTTTAAAAAA GAGCTTTAGA   383
3civ-1      179 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   228
4cvii-1      95 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   144
3cx-3        45 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA    94
3ci          82 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   131
4av         105 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   154
2cii        103 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   152
3cix         61 CTAAAGCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   110
4avii        70 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   119
2ciii-4     104 CTCTTCCAGA CATCGAAGCC TCAGGAAAGT GCTTAAAAAG AGCTTAAGAA   153
```

Figure 17. Alignment of the 3' sequences of various castor cDNA clones isolated with pCRS677. Note that the linker sequences (CTCTAAAG) have not been removed. Similarity between 3cvii and the other clones begins at position 313 of this alignment.

It was concluded that the first 305 bp of the cDNA in clone 3cvii contained extraneous DNA, not related to pCRS677 (nor any other known sequence). Further sequence data was obtained only from clone 3civ-1, hereafter designated pFL2.

Sequencing primers were designed from the assembled contiguous sequence, and used to sequence both strands of the entire clone pFL2 (Figure 18). The clone encodes a 186 bp 5' untranslated region (i.e. before the first ATG codon), an 1161 bp open reading frame, and a 101 bp 3' untranslated region, including a short (9 bp) poly(A) tail. The open reading frame encodes a 387 amino acid protein with a predicted molecular weight of 44406.8. The amino terminus lacks features of a typical signal peptide.[11] The predicted sequence of the *Brassica napus fad3* microsomal desaturase also lacks a typical signal peptide.[12]

Comparison of the pFL2 nucleotide and deduced amino acid sequences with sequences of membrane-bound desaturases (Table 13) indicates that pFL2 is homologous to these genes. *Arabidopsis fad6* and *fad8* sequences used for comparison in Table 13 were unpublished data of S. Gibson and V. Arondel. Sequence similarity between pFL2 and these desaturase genes is considerably weaker than similarities among the desaturase genes (see for example Figure 10, chapter 4). The most similar sequence to pFL2 is that of a desaturase-homologue of unknown function cloned from *Petroselinum crispum* (parsley) in the laboratory of Dr. K. Hahlbrock (I. Sommsich, personal communication). An alignment of the deduced amino acid sequences of pFL2, *Brassica napus* fad3 and pFL1 (the castor *fad7* cDNA isolated in chapter 4, above) is shown in Figure 19. The sequence of *fad2*, the

Figure 18. DNA and deduced amino acid sequence of the clone pFL2. Motifs conserved among membrane-bound desaturases are shaded.

```
   1  GCC ACC TTA AGC GAG CGC CGC ACA CGA AGC CTC CTT TCA CAC TTG GTG ACC TCA AAT CAA

  61  ACA CCA CAC CTT ATA ACT TAG TCT TAA GAG AGA GAG AGA GAG AGA GGA GAC ATT TCT CTT

 121  CTC TGA GAT AAG CAC TTC TCT TCC AGA CAT CGA AGC CTC AGG AAA GTG CTT AAA AAG AGC

                M   G   G   G   G   R   M   S   T   V   I   T   S   N   N   S   E   K
 181  TTA AGA ATG GGA GGT GGT GGT CGC ATG TCT ACT GTC ATA ACC AGC AAC AAC AGT GAG AAG

  19  K   G   G   S   S   H   L   K   R   A   P   H   T   K   P   P   F   T   L   G
 241  AAA GGA GGA AGC AGC CAC CTT AAG CGA GCG CCG CAC ACG AAG CCT CCT TTC ACA CTT GGT

  39  D   L   K   R   A   I   P   P   H   C   F   E   R   S   F   V   R   S   F   S
 301  GAC CTC AAG AGA GCC ATC CCA CCC CAT TGC TTT GAA CGC TCT TTT GTG CGC TCA TTC TCC

  59  Y   V   A   Y   D   V   C   L   S   F   L   F   Y   S   I   A   T   N   F   F
 361  TAT GTT GCC TAT GAT GTC TGC TTA AGT TTT CTT TTC TAC TCG ATC GCC ACC AAC TTC TTC

  79  P   Y   I   S   S   P   L   S   Y   V   A   W   L   V   Y   W   L   F   Q   G
 421  CCT TAC ATC TCT TCT CCG CTC TCG TAT GTC GCT TGG CTG GTT TAC TGG CTC TTC CAA GGC

  99  C   I   L   T   G   L   W   V   I   G   H   E   C   G   H   H   A   F   S   E
 481  TGC ATT CTC ACT GGT CTT TGG GTC ATC GGC CAT GAA TGT GGC CAT CAT GCT TTT AGT GAG

 119  Y   Q   L   A   D   D   I   V   G   L   I   V   H   S   A   L   L   V   P   Y
 541  TAT CAG CTG GCT GAT GAC ATT GTT GGC CTA ATT GTC CAT TCT GCA CTT CTG GTT CCA TAT

 139  F   S   W   K   Y   S   H   R   R   H   H   S   N   I   G   S   L   E   R   D
 601  TTT TCA TGG AAA TAT AGC CAT CGC CGC CAC CAT TCT AAC ATA GGA TCT CTC GAG CGA GAC

 159  E   V   F   V   P   K   S   K   S   K   I   S   W   Y   S   K   Y   S   N   N
 661  GAA GTG TTC GTC CCG AAA TCA AAG TCG AAA ATT TCA TGG TAT TCT AAG TAC TCA AAC AAC

 179  P   P   G   R   V   L   T   L   A   A   T   L   L   L   G   W   P   L   Y   L
 721  CCG CCA GGT CGA GTT TTG ACA CTT GCT GCC ACG CTC CTC CTT GGC TGG CCT TTA TAC TTA

 199  A   F   N   V   S   G   R   P   Y   D   R   F   A   C   H   Y   D   P   Y   G
 781  GCT TTC AAT GTC TCT GGT AGA CCT TAC GAT CGC TTT GCT TGC CAT TAT GAT CCC TAT GGC

 219  P   I   F   S   E   R   E   R   L   Q   I   Y   I   A   D   L   G   I   F   A
 841  CCA ATA TTT TCC GAA AGA GAA AGG CTT CAG ATT TAC ATT GCT GAC CTC GGA ATC TTT GCC

 239  T   T   F   V   L   Y   Q   A   T   M   A   K   G   L   A   W   V   M   R   I
 901  ACA ACG TTT GTG CTT TAT CAG GCT ACA ATG GCA AAA GGG TTG GCT TGG GTA ATG CGT ATC

 259  Y   G   V   P   L   L   I   V   N   C   F   L   V   M   I   T   Y   L   Q   H
 961  TAT GGG GTG CCA TTG CTT ATT GTT AAC TGT TTC CTT GTT ATG ATC ACA TAC TTG CAG CAC

 279  T   H   P   A   I   P   R   Y   G   S   S   E   W   D   W   L   R   G   A   M
1021  ACT CAC CCA GCT ATT CCA CGC TAT GGC TCA TCG GAA TGG GAT TGG CTC CGG GGA GCA ATG

 299  V   T   V   D   R   D   Y   G   V   L   N   K   V   F   H   N   I   A   D   T
1081  GTG ACT GTC GAT AGA GAT TAT GGG GTG TTG AAT AAA GTA TTC CAT AAC ATT GCA GAC ACT

 319  H   V   A   H   H   L   F   A   T   V   P   H   Y   H   A   M   E   A   T   K
1141  CAT GTA GCT CAT CAT CTC TTT GCT ACA GTG CCA CAT TAC CAT GCA ATG GAG GCC ACT AAA

 339  A   I   K   P   I   M   G   E   Y   Y   R   Y   D   G   T   P   F   Y   K   A
1201  GCA ATC AAG CCT ATA ATG GGT GAG TAT TAC CGG TAT GAT GGT ACC CCA TTT TAC AAG GCA

 359  L   W   R   E   A   K   E   C   L   F   V   E   P   D   E   G   A   P   T   Q
1261  TTG TGG AGG GAG GCA AAG GAG TGC TTG TTC GTC GAG CCA GAT GAA GGA GCT CCT ACA CAA

 379  G   V   F   W   Y   R   N   K   Y   *
1321  GGC GTT TTC TGG TAC CGG AAC AAG TAT TAA AAA AGT GTC ATG TAG CCT GTT TCT TTA AGA

1381  GAA GTA ATT AGA ACA AGA AGG AAT GTG TGT GTA GTG TAA TGT GTT CTA ATA AAG AAG GCA

1441  AAA AAA AA
```

Table 13. Sequence similarity at the nucleotide (NT) and amino acid (AA) levels between pFL2 and membrane-bound desaturase genes. Nucleotide sequence of the clone ELI72 was not available for comparison, and the deduced amino acid sequence used here is considered preliminary data only.

			Sequence Similarity	
Gene	**Organism**	**Ref.**	**NT**	**AA**
ELI72	*Petroselinum crispum*	text		56
fad7	*Ricinus communis*	chap 4	47.1	38.6
fad3	*Brassica napus*	12	46.5	37.4
fad8	*Arabidopsis thaliana*	text	46.5	36.2
fad7	*Arabidopsis thaliana*	13	47.4	35.5
fad6	*Arabidopsis thaliana*	text	46.5	23.5
desA	*Synechoccocus* sp.	14	45.5	22.6

```
                10         20         30         40         50
pFL2      -75 .......... .......... .......... .......... .......... -26
Fad3      -79 .......... .......... .......... .......... .......... -30
pFL1        1 MAAGWVLSEC GLRPLPRIYS RPRIGFTSKT TNLLKLRELP DSKSYNLCSS  50

                60         70         80         90        100
pFL2      -25 .......... .......... .....MGGGG RMSTVITSNN SEKKGGSSHL  25
Fad3      -29 .......... .......... .........M VVAMDQR**V NGDS*ARKEE  21
pFL1       51 FKVSSWSNSK QSNWALNVAV PVNVSTVS*E DDREREEF*G IVNVDEGKGE 100

               110        120        130        140        150
pFL2       26 KRAPHTKPPF TLGDLKRAIP PHCFERSFVR SFSYVAYDVC LSFLFYSIAT  75
Fad3       22 GFD*SAQ*** KI**IRA*** K**WVK*PL* *M***TR*I- FAVAALAM*A  71
pFL1      101 FFDAGAP*** **A*IRA*** K**WVKNPW* *M***LR**V VV*GLAAV*- 150

               160        170        180        190        200
pFL2       76 NFFPYISSPL SYVAWLVYWL FQGCILTGLW VIGHECGHHA FSEYQLADDI 125
Fad3       72 VY*------D *WFL*PL**V A**TLFWAIF *L**D***GS **DIP*LNSV 121
pFL1      151 ---A*FNN-- -W***PL**F C**TMFWA*F *L**D***GS **NNPKLNSV 200

               210        220        230        240        250
pFL2      126 VGLIVHSALL VPYFSWKYSH RRHHSNIGSL ERDEVFVPKS KSKISWYSKY 175
Fad3      122 **H*L**FI* ***HG*RI** *T**Q*H*HV *N**SW**LP -E*L--*KNL 171
pFL1      201 **HLL**SI* ***HG*RI** *T**Q*H*HV *N**SWH*L* -E**--FKSL 250

               260        270        280        290        300
pFL2      176 SNNPPGRVLT LAATL-LLGW PLYLAFNVSG RPYDRFACHY DPYGPIFSER 225
Fad3      172 P--HST*M*R YTVP*PM*AY *I**WYRSP* ----KEGS*F N**SSL*APS 221
pFL1      251 D*VTKTLRFS *PFP--M*AY *F**----WS *SPGKKGS*F H*DSGL*VPK 300

               310        320        330        340        350
pFL2      226 ERLQIYIADL --GIFATTFV LYQATMAKGL AWVMRIYGVP LLIVNCFLVM 275
Fad3      222 **KL*ATSTT CWS*MLA*L* -*LSFLVDPV T-*LKV**** YI*FVMW*DA 271
pFL1      301 **KD*ITSTA --CWT*MAAL *VYLNFSM*P VQMLKL**I* YW*FVMW*DF 350

               360        370        380        390        400
pFL2      276 ITYL-QHTHP A-IPRYGSSE WDWLRGAMVT VDRDYGVLNK VFHNIADTHV 325
Fad3      272 V***HH*G*D EKL*W*RGK* *SY***GLT* I*****IF*N IH*D*-G*** 321
pFL1      351 V***HH*G*E DKL*W*RGKA *SY***GLT* L*****WI*N IH*D*-G*** 400

               410        420        430        440        450
pFL2      326 AHHLFATVPH YHAMEATKAI KPIMGEYYRY DGTPFYKALW REAKECLFVE 375
Fad3      322 I****PQI** **LVD**R*A *HVL*R***E PK*SGAIPIH LVESLVASIK 371
pFL1      401 I****PQI** **LV***E*A **V**K***E PKKSGPLP*H LLGSLVRSMK 450

               460        470        480        490        500
pFL2      376 PDEGAPTQGV FWYRNKY... .......... .......... .......... 425
Fad3      372 K*HYVSDT*D IVFYETDPDL YVYASDKSKI N......... .......... 421
pFL1      451 E*HYVSDT*D VV*YQ*DPKL SGIGGEKTE. .......... .......... 500
```

Figure 19. Alignment of the deduced amino acid sequence of pFL2 with pFL1 (Fad7 from castor: chapter 4), and *Brassica napus* Fad3.[12]

desaturase gene hypothesized to be most similar to the oleate-12-hydroxylase gene, was not available for comparison.

The deduced amino acid sequence of pFL2 (Figures 18, 16) contains the conserved histidine-rich repeats (HXXHH) also found in other desaturases (chapter 4). These motifs are separated by 169 intervening amino acid residues. The motifs are separated by only 122 residues in the deduced amino acid sequence of the *Arabidopsis thaliana fad2* gene (John Browse, personal communication to Sue Gibson). This information suggests that pFL2 does not encode the microsomal Δ12 desaturase. A divergent form (GHECGH) of the conserved desaturase motif (GHDCGH) used for priming PCR reactions (chapter 4), was found in the pFL2 sequence. Since the codons for glutamic acid (E) differ from the codons for aspartic acid (D) only at the third position, priming in the PCR reaction would have resulted from only a single-base mis-match. Despite attempts to select cycling conditions (viz. annealing temperature) at which some mis-match could occur, the clones analysed in chapter 4 almost all contained the authentic primer sequence, indicating that mis-match did not occur, and hence the pFL2 sequence was not amplified.

Expression of pFL2 in Castor

Since oleate-12-hydroxylase activity is only found in the developing seeds of castor,[1-4] it was of great interest to determine whether pFL2 is also expressed only in seeds, or is also expressed in other tissues. This question was addressed by testing

for hybridization of pFL2 to RNA purified from developing seeds and from leaves.

A northern blot of RNA from leaves and developing seeds (stage III to stage V[15]) of castor was probed with the ^{32}P-labelled insert of clone pCRS677, which corresponds to ~700 bp of the 3' end of pFL2. Brief (30 min) exposure of the blot to X-ray film revealed that the probe hybridised to a single band of ~1.67 kb, only in the seed RNA lane (Figure 20, panel A). Upon overexposure (16 h) of the film, a band of similar size was detected in the leaf RNA lane, in addition to a second, larger, band in the seed RNA lane (Figure 20, panel B). The blot was also exposed to a phosphor-imaging screen, for quantitation of probe hybridisation. Total exposure to this screen in an area covering the band in the leaf lane was 4.36×10^4 units above background. Total exposure in an area of equal size over the major band in the seed lane was 1.17×10^7 units above background, 268-fold more than in the leaf lane. The blot was re-probed with a β-tubulin gene, which gave bands of equal intensity in the seed and leaf lanes (Figure 20, panel C), verifying that equal quantities of undegraded RNA were loaded in the two lanes.

These results show that pFL2 is highly and specifically expressed in seed of castor. Over-exposure of the northern revealed a 268-fold weaker band of similar size in leaf RNA, but also a second band in seed RNA, suggesting that these bands are due to weak hybridization of pFL2 to related sequences, such as desaturases. Strong seed-specific expression of pFL2 is compatible with the possibility that it encodes oleate-12-hydroxylase. No desaturase or other putative desaturase-homologue which would have these characteristics is known to be expressed in castor.

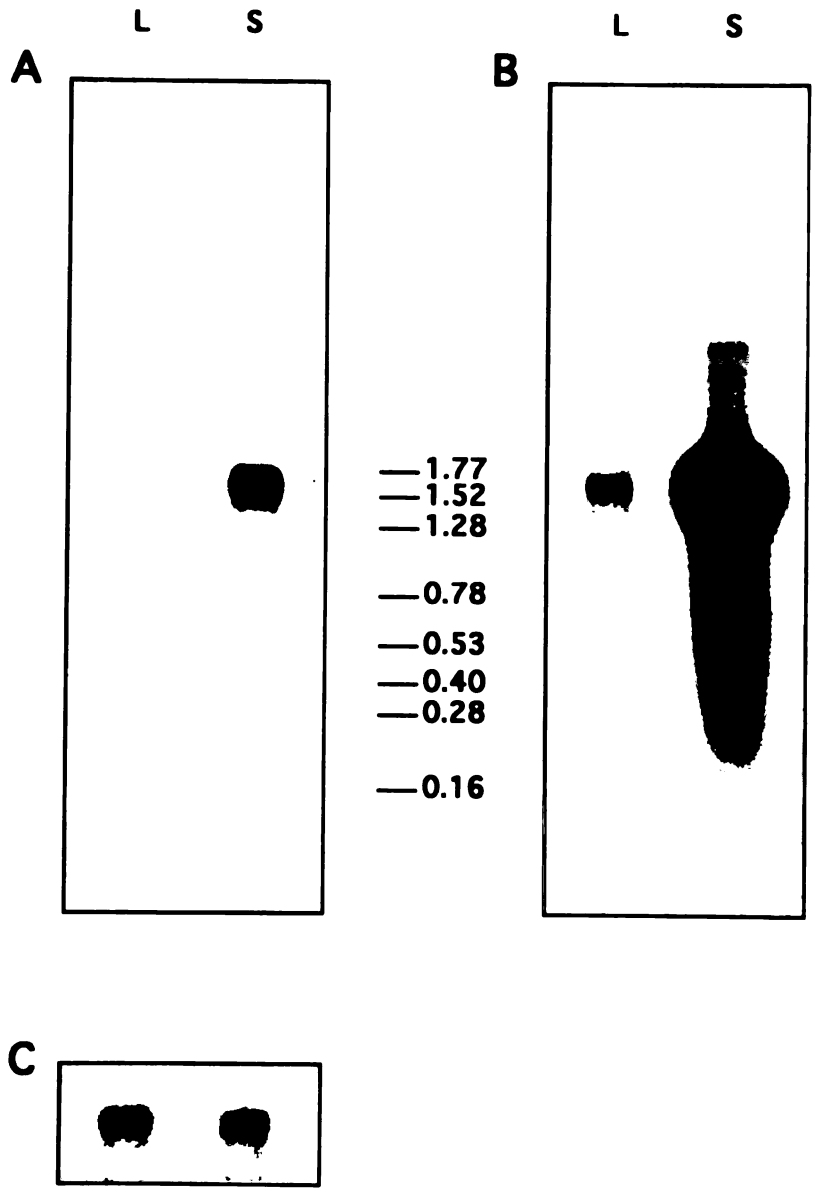


Figure 20. Northern blot analysis of pFL2 expression in castor. A ^{32}P-labelled probe corresponding to ~700 bp of the 3' end of clone pFL2 was hybridised to poly(A)$^+$ RNA from leaves (L) and developing seeds (S) of castor. Panel A: the blot was exposed to film for 30 min. The migration of RNA standards (kb) is shown to the right. Panel B: the same blot was exposed for 16 h. Panel C: the same blot was hybridised to a ^{32}P-labelled probe made from the *Colletotrichum graminicola* β-tubulin gene *TUB2*.[6]

Southern Analysis with pFL2

Southern analysis was used to examine the copy number of genes in the castor genome corresponding to clone pFL2, and to examine whether related sequences could be detected in the castor genome, and in the genome of a different plant, in which oleate-12-hydroxylase is absent.

Genomic DNAs of castor and of *Arabidopsis thaliana* were digested with restriction enzymes, separated by agarose gel electrophoresis and transferred to a nylon membrane. *Arabidopsis* was chosen for the negative control DNA because it has no known oleate-12-hydroxylase, and the DNA was readily available. The membrane was hybridised with the ^{32}P-labelled insert of clone pFL2 at 65°C, and exposed to X-ray film. The probe hybridised with a single band in each digest of castor DNA, but did not hybridise to the *Arabidopsis* DNA (Figure 21), indicating that the gene from which pFL2 was transcribed is present in a single copy in the castor genome, and is not present in the *Arabidopsis* genome. The blot was then hybridised again, with an identical probe, but at less stringent hybridization conditions (52°C) (Figure 22). This revealed additional weakly-hybridising bands. In castor DNA, a total of four bands were detected in both the *Eco*RI digest and the *Bam*HI digest. In *Arabidopsis* DNA, four bands (*Eco*RI), five bands (*Bam*HI), or possibly three bands (*Hin*dIII) were detected. These results suggest that two or more genes with sequence similarity to pFL2 occur in both the castor and *Arabidopsis* genomes, in agreement with sequence similarity between pFL2 and desaturase genes (above).

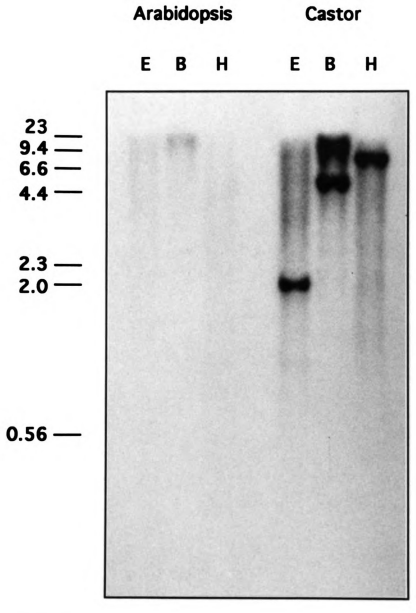


Figure 21. A Southern blot of genomic DNA from *Arabidopsis thaliana* and castor (*Ricinus communis*) digested with restriction enzymes *Eco*RI (E), *Bam*HI (B), or *Hind*III (H), was hybridised at high stringency (65°C) with the ^{32}P-labelled insert of clone pFL2. Migration of DNA standards (kb) is shown to the left.

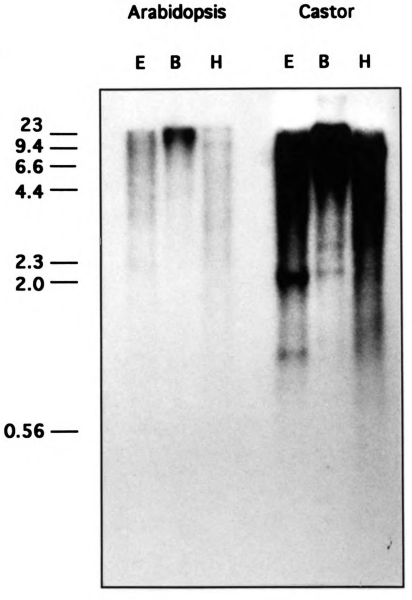


Figure 22. The Southern blot shown in Figure 21 was hybridised again with the pFL2 probe under low stringency (52°C). (See legend to Figure 21 for details).

Expression of pFL2 in Transgenic Yeast

The experiments described above indicate that pFL2 has features expected of an oleate-12-hydroxylase gene, but do not provide convincing evidence that pFL2 is in fact the oleate-12-hydroxylase gene. The function of the protein encoded by pFL2 was, therefore, examined by expression in transgenic cells. Since the isolation of clones using pCRS677 as a probe, described above, made use of the pYES2.0 library, designed for expression of the cloned cDNA in yeast (chapter 3), it was a simple experiment to express pFL2 in transgenic yeast.

In addition to pFL2, two other clones, 2cii and 3ci, were transformed into yeast cells by electroporation. Sequence data and digestion with the enzyme *Kpn*I (which cuts near the 3' end of pFL2 and also in the cloning site of the vector directly downstream of the Gal 1 promoter (Figure 5), and is therefore diagnostic for the orientation of the insert DNA) indicated that 2cii and 3ci contained the full coding sequence, but were in the reverse orientation for expression in pYES2.0, while pFL2 was in the forward orientation. Clones 2cii and 3ci could therefore be used as negative controls. Two yeast strains were used in the experiment, CGY2557 (chapter 3) and a protease-deficient strain, JO522. Transformants were patched onto selective medium (lacking uracil) with either glucose or galactose as carbon source. Under the control of the GAL promoter, the castor DNA would only be transcribed in cells grown on the galactose-containing medium. Fatty acid methyl esters (FAMEs) prepared from the harvested yeast cells were analysed by gas chromatography.

FAMEs from each of the three clones grown on both carbon sources were analysed for the host strain CGY2557, and FAMEs from duplicate patches for each of these combinations were analysed for the host strain JO522. Growth of yeast harbouring pFL2 on medium containing galactose did not result in the appearance of any material chromatographing at the position of authentic methyl-ricinoleate. Furthermore, no novel peaks were detected in any of the chromatograms. A typical chromatogram is shown in Figure 23, with the authentic methyl-ricinoleate standard for comparison. Expression of the castor clones also caused no changes in the relative proportions of yeast FAMEs, though it should be noted that the unsaturated 18-carbon fatty acids were poorly resolved on the column used in this experiment.

The results of this experiment neither support nor contradict the possibility that pFL2 encodes oleate-12-hydroxylase. If pFL2 does encode oleate-12-hydroxylase, the lack of expression observed in this experiment indicates a possible reason why pFL2 was not isolated in the experiments reported in chapter 3.

Expression of pFL2 in Transgenic Plants

The insert of clone pFL2 was ligated between the 35S promoter and nos terminator of the plant expression binary vector pBI121,[7] in the correct orientation for expression of the open reading frame, by two independent cloning strategies. These constructs were transformed into *Agrobacterium tumefaciens* strains R1000, GV3101, and LBA4404. R1000 is a C58-derivative cured of the Ti plasmid and harbouring the

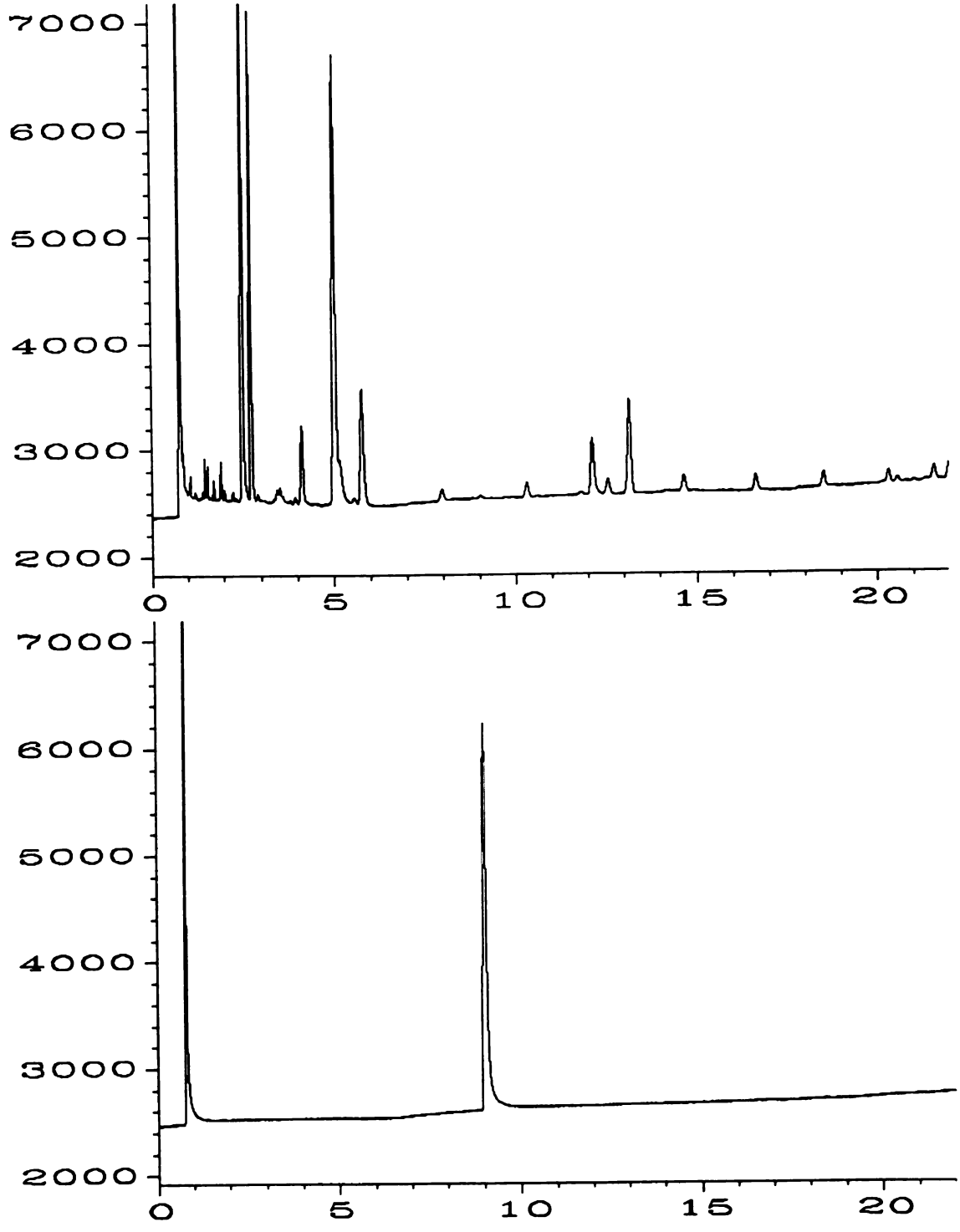


Figure 23. Typical gas chromatogram (above) of yeast fatty acid methyl esters (CGY2557 harbouring pFL2, grown on glucose medium), and chromatogram of methyl-ricinoleate standard (below). Detector signal is plotted against retention time (min).

A. rhizogenes plasmid pRiA4b,[16] and was used to produce transgenic roots on carrot disks. GV3101[17] and LBA4404[18] contain disarmed Ti plasmids, and were used to transform both tobacco leaf explants and tobacco suspension culture cells. Transgenic tissues or callus were analysed by gas chromatography for fatty acid composition in three experiments.

Experiment 1. Roots produced on the inoculated carrot disks were excised and tested for growth in medium containing kanamycin (50 mg l^{-1}). Transformation of carrot roots using the strain R1000 involves cotransformation of the binary vector and the Ri plasmid. Seven of the eight roots analysed in this experiment were kanamycin resistant, indicating they were cotransformed. These represented construct B6 (roots 8-1, 11-1, 22-1), construct 9/18 3 (roots 20-1, 20-2, 20-3), and the intact vector pBI121 (root 23-1).

Root FAMEs were treated with silylating reagent and analysed by gas chromatography using a mass-selective detector (electron impact-mass spectrometry). No peaks were observed in the chromatograms at the elution time of authentic trimethylsilyloxy-methyl-ricinoleate. A typical chromatogram is shown in Figure 24, and a chromatogram of the standard is shown in Figure 25. Examination of the ions present in the background signal in this part of the chromatogram failed to detect the major ion (m/z 187) produced from the authentic standard. Mass spectra taken at similar retention times are shown in Figure 26 (corresponding to the chromatogram shown in Figure 24) and Figure 27 (corresponding to the standard, Figure 25).

Figure 24. Typical gas chromatogram (experiment 1) of transformed carrot root TMS-derivatised fatty acid methyl esters (root 20-1). Detector signal is plotted against retention time (min). Peaks identified by co-chromatography with standards, and by comparison of mass spectra to a mass-spectral data library: 17.38 min, 16:0; 20.88 min, 18:2; 21.04 min and 21.19 min, 18:1; 21.69 min, 18:0.

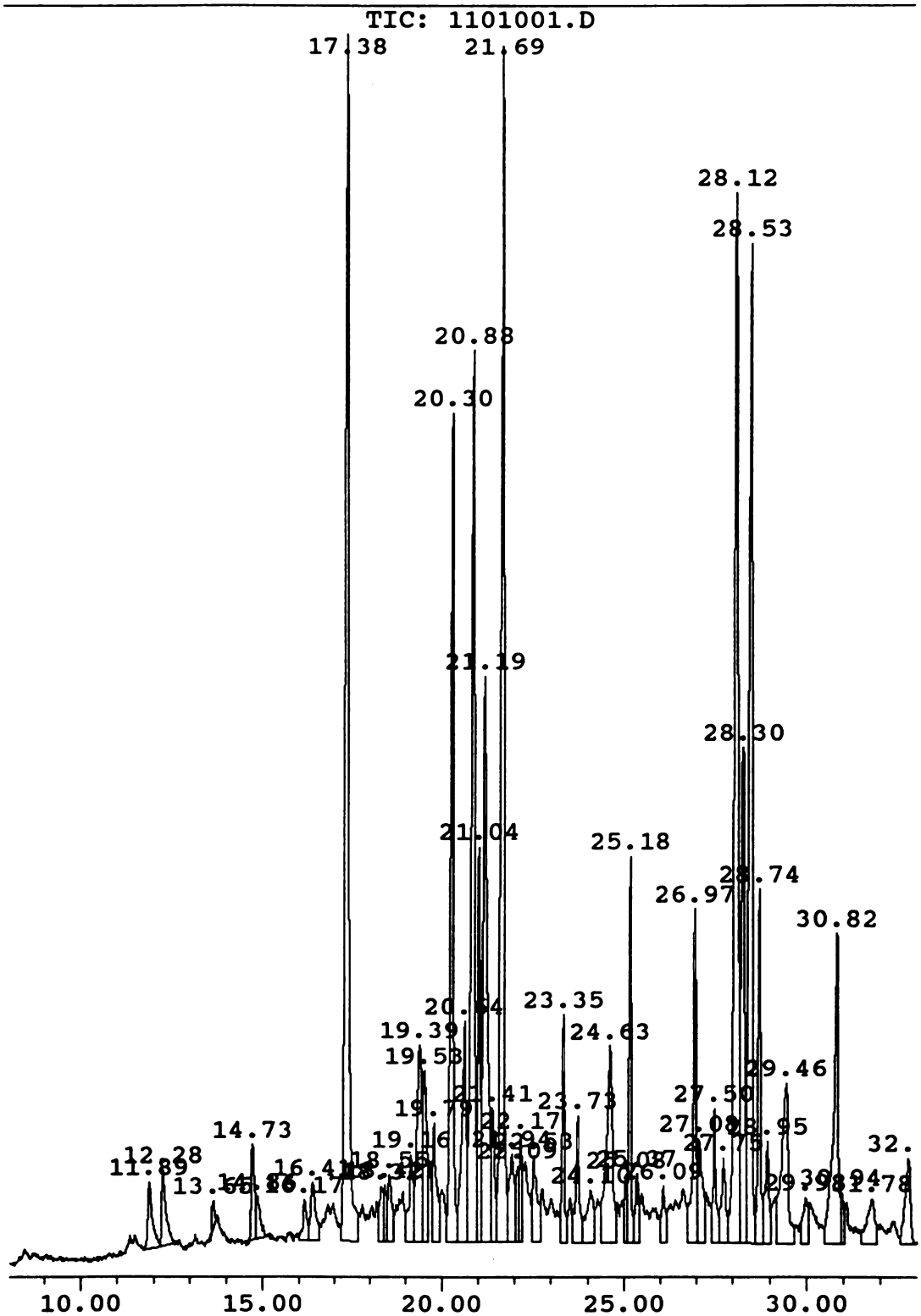

TIC: 1101001.D
17.38
21.69
28.12
28.53
20.88
20.30
21.19
28.30
21.04
25.18
28.74
26.97
30.82
23.35
20.64
19.39
24.63
19.53
29.46
27.50
23.73
14.73
12.28
11.89
10.00
15.00
20.00
25.00
30.00

Figure 25. Gas chromatogram of trimethylsilyloxy-methyl-ricinoleate (experiment 1). Detector signal is plotted against retention time (min). Retention time of standard: 25.1 min.

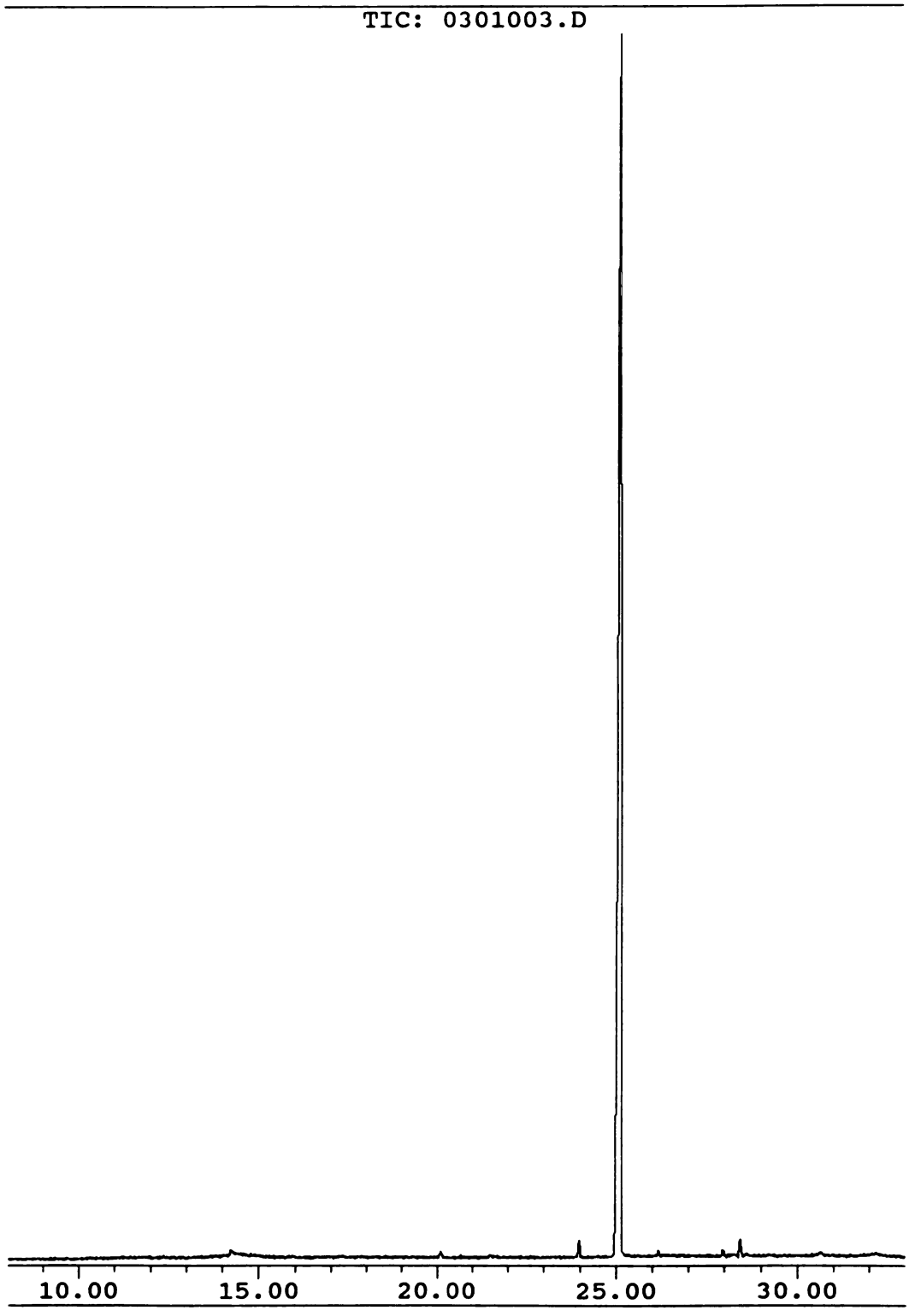

TIC: 0301003.D
10.00
15.00
20.00
25.00
30.00

Figure 26. Mass-spectrum of ions derived from material eluting from the column at 25.165 min, in the chromatogram shown in Figure 24.

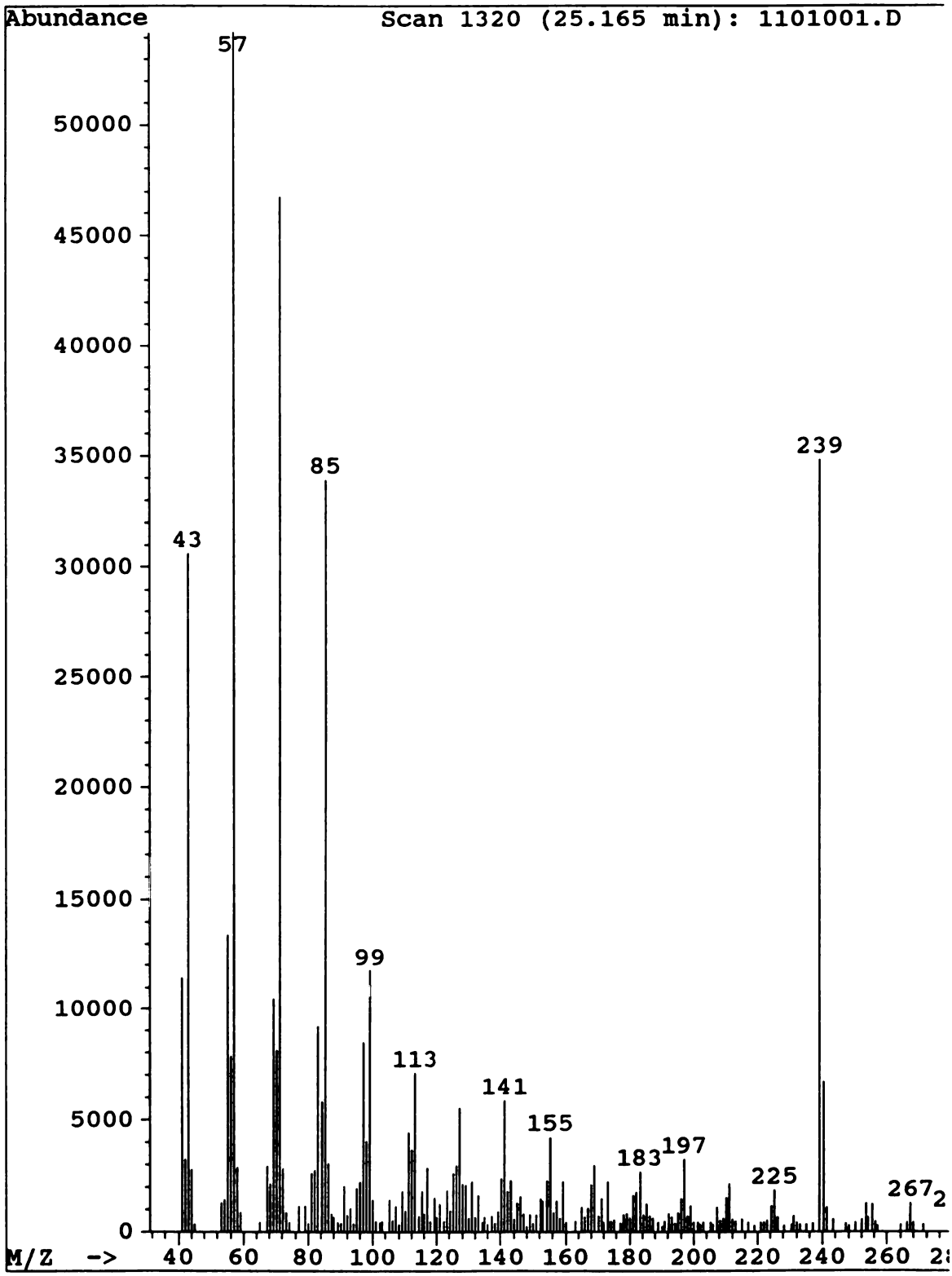
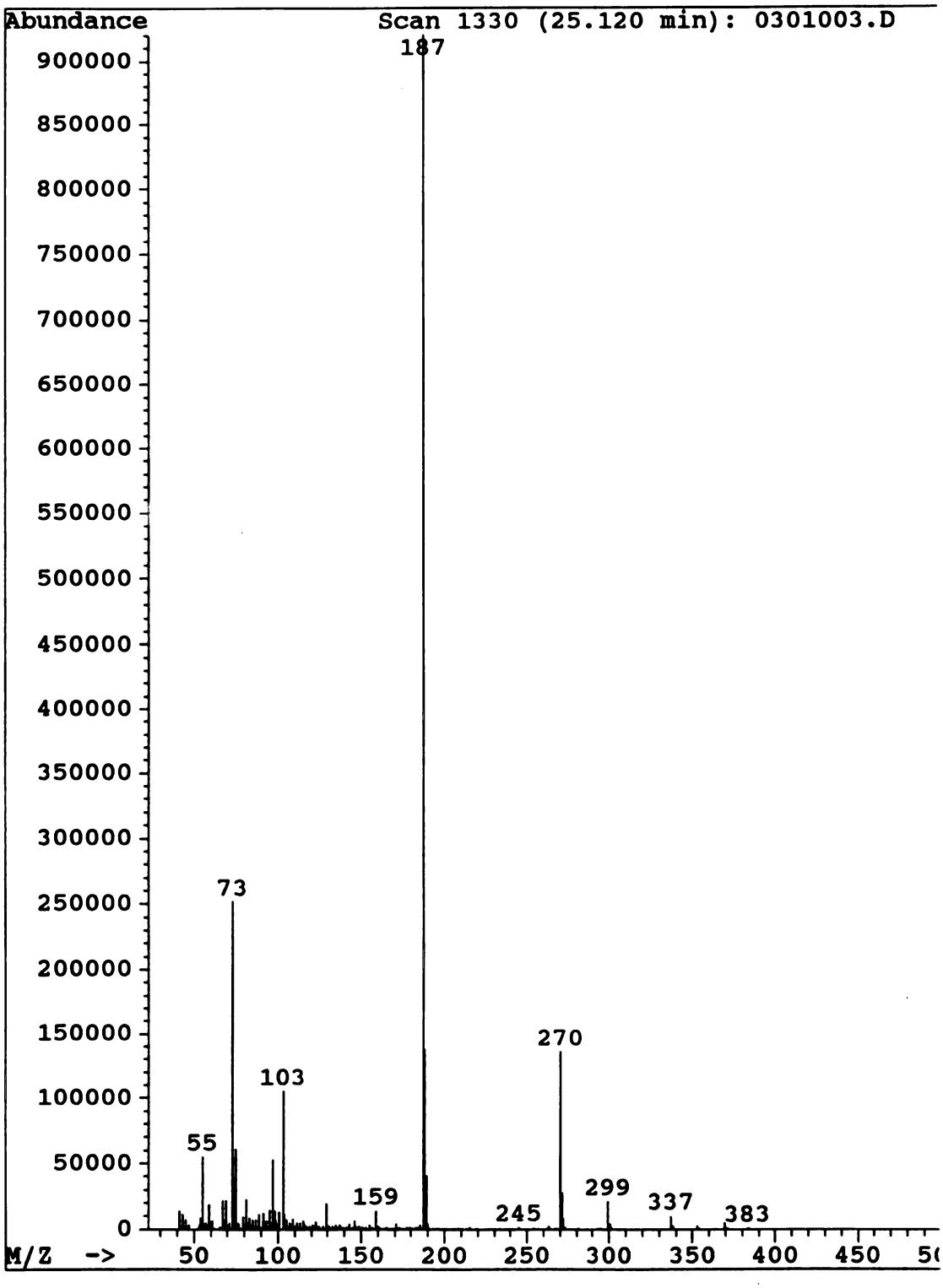

Abundance
Scan 1320 (25.165 min): 1101001.D
57
50000
45000
40000
35000
85
239
43
30000
25000
20000
15000
99
10000
113
141
5000
155
183
197
225
267
0
M/Z ->
40 60 80 100 120 140 160 180 200 220 240 260

Figure 27. Mass-spectrum of trimethylsilyloxy-methyl-ricinoleate, taken from the chromatogram shown in Figure 25 at 25.120 min.


Abundance
Scan 1330 (25.120 min): 0301003.D
187
900000
850000
800000
750000
700000
650000
600000
550000
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
73
103
55
159
245
270
299
337
383
M/Z ->
50
100
150
200
250
300
350
400
450

The mass spectrum of TMS-methyl-ricinoleate shown in Figure 27 is the same as that previously established for this compound (Appendix A, which includes an explanation of the fragmentation pattern resulting in the ions observed). It was concluded that the samples contained no ricinoleic acid. The relative content of the abundant fatty acids in each sample is presented in Table 14. Octadecadienoic and octadecatrienoic acids were not resolved on the column used in this experiment. Though there was some variability in the combined content of these particular fatty acids, there appeared to be no correlation with expression of pFL2.

Experiment 2. Transformed carrot roots which were growing in medium containing 50 mg l^{-1} kanamycin were analysed for fatty acid composition. Eighteen independently-transformed roots were analysed, including the seven roots previously analysed in experiment 1. These represented the four constructs used for transformation: 9/18 3 (5 roots), B6 (7), A4 (1) and pBI121 (5). Also analysed was shoot tissue from nine transformed tobacco shoots derived from the tobacco leaf explant transformation. Eight of these nine shoots subsequently rooted in medium containing 100 mg l^{-1} kanamycin; it is not certain that the ninth shoot is truly transformed.

FAMEs from these samples were first analysed by gas chromatography without TMS derivatization, since the TMS reagent elutes as a broad peak which comigrates approximately with stearic acid. After the relative composition of usual fatty acids had been examined, samples were derivatised with TMS and re-injected.

Table 14. Relative content (%) of the abundant fatty acids in transgenic carrot roots. Fatty acid methyl ester abbreviations: 16:0 hexadecenoic acid, 18:2 octadecadienoic acid, 18:3 octadecatrienoic acid, 18:1 octadecenoic acid (two isomers were resolved but not identified; carrot contains petroselenic acid, Δ6-18:1, in addition to oleic acid, Δ9-18:1), 18:0 stearic acid.

			Fatty Acid Methyl Ester				
			16:0	18:2 +18:3	18:1	18:1	18:0
		Retention Time (min)	17.28	20.81	20.97	21.11	21.60
constr.	kan^R	root					
B6	+	A8-1	28.2	7.0	9.9	23.9	26.7
B6	+	A11-1	30.1	3.8	10.0	25.0	27.5
B6	+	22-1	30.5	22.1	7.9	17.6	19.6
9/18 3	+	20-1	28.7	20.7	6.9	16.5	24.4
9/18 3	+	20-2	29.3	15.5	7.9	18.5	25.7
9/18 3	+	20-3	30.1	7.4	8.8	21.0	29.4
pBI121	+	23-1	29.4	9.5	9.1	22.5	26.3
pBI121	-	A3-1	27.9	12.0	9.2	22.7	24.6

The relative content of the abundant fatty acids in each sample is presented in Table 15. Considerable variation was again observed in the relative contents of octadecadienoic and octadectrienoic acids, but there was no clear correlation with expression of pFL2. It appears that samples may differ considerably from each other simply due to differences in the age of the tissue, its physiological state, or particular environment. When the samples were re-injected following TMS derivatization, no peaks were found in the position of authentic trimethylsilyloxy-methyl-ricinoleate in any of the chromatograms.

Experiment 3. A gene involved in the biosynthesis of one other unusual fatty acid, petroselinic acid, has been cloned.[19] This gene encodes an acyl-ACP desaturase, and resulted in the production of petroselinic acid and hexadec-4-enoic acid when expressed in transgenic tobacco callus. These unusual fatty acids, normally absent from tobacco, comprised between 1% and 4% of total fatty acids in the transgenic samples.[19] However, differentiated plant tissues regenerated from these transgenic calli, contained little or no detectable levels of these unusual fatty acids (E. Cahoon, personal communication).

In case the results of experiments 1 and 2, above, were due to a similar phenomenon whereby no detectable ricinoleic acid accumulated in the transgenic carrot root or tobacco shoot tissue, this experiment focused upon analysis of transgenic tobacco calli. Calli from both the tobacco leaf explant transformation

Table 15. Relative content (%) of the abundant fatty acids in transgenic carrot roots and tobacco shoots. Fatty acid methyl ester abbreviations: 16:0 hexadecenoic acid, 18:0 stearic acid, 18:1 octadecenoic acid, 18:2 octadecadienoic acid, 18:3 octadecatrienoic acid.

			Fatty Acid Methyl Ester				
			16:0	18:0	18:1	18:2	18:3
	Retention Time (min)		7.16	10.12	10.83	12.06	13.55
			Tobacco				
constr.	kan^R	shoot					
B6	+	10-3	26.9	9.8	6.4	14.1	38.8
B6	+	10-1	31.1	7.4	8.9	13.3	36.3
B6	+	6-1	31.8	5.4	1.2	22.0	38.0
B6	+	10-2	28.9	6.5	1.9	16.6	44.4
A4	+	9-3	32.9	4.9	3.0	24.8	32.8
A4	+	9-1	28.9	7.4	4.3	18.7	37.1
A4	-	9-2	28.8	8.7	4.7	10.1	44.9
9/18 3	+	8	29.9	8.5	4.7	9.9	43.0
pBI121	+	4/12-1	29.2	7.5	4.7	11.7	43.6
			Carrot				
constr.	kan^R	root					
B6	+	18-5	34.2	9.2	15.5	22.4	14.4
B6	+	18-4	33.4	9.8	18.5	17.3	15.2
B6	+	A11-6	33.2	18.0	22.5	1.5	17.0
B6	+	A11-4	31.9	19.0	18.4	3.5	17.8
B6	+	22-1	35.8	18.0	18.7	9.9	14.6
B6	+	A11-1	33.4	18.2	21.4	5.4	18.0
B6	+	A8-1	36.4	11.8	11.5	24.6	12.0
A4	+	A4-4	29.8	15.5	19.3	7.6	21.4
9/18 3	+	20-1	30.9	11.7	16.2	18.3	16.0
9/18 3	+	20-5	32.8	19.1	19.0	5.0	18.1
9/18 3	+	20-10	34.3	11.3	14.6	21.4	13.7
9/18 3	+	20-3	34.6	11.7	14.0	20.9	14.7
9/18 3	+	20-2	34.4	8.7	10.4	30.6	13.1
pBI121	+	19-4	35.2	11.4	18.8	16.9	13.0
pBI121	+	23-6	33.7	17.5	22.6	1.6	19.2
pBI121	+	11-1	30.3	14.8	22.3	2.5	20.5
pBI121	+	23-5	33.4	16.0	21.5	3.7	17.9
pBI121	+	23-1	34.9	18.0	17.4	10.7	15.1

system and the tobacco suspension culture (NT-1) cell transformation system were analysed in this experiment. Some additional transgenic tobacco shoot samples were also analysed. In the tobacco leaf explant transformation system, callus may form at the margins of the explants, in addition to callus formed at the points of inoculation (by stabbing the leaf with a needle bearing *Agrobacterium* cells). Only callus clearly arising from these inoculation sites was analysed in this experiment, and only shoots derived from such calli were considered transgenic.

FAMEs were prepared from 7 calli and 7 shoots (2 of which included some callus tissue in the sample analysed, and all of which had rooted--or subsequently rooted--in kanamycin-containing medium) from the leaf explant system, and 4 calli from the NT-1 system. In addition, samples of each type were analysed that had been transformed with the intact pBI121 vector, as negative controls.

The relative content of the abundant fatty acids in each sample is presented in Table 16. As in the previous experiments, considerable variation was seen in the relative proportions particularly of octadecadienoic and octadecatrienoic acids, but this did not correlate with expression of pFL2. When the samples were re-injected following TMS derivatization, no peaks were found in the position of authentic trimethylsilyloxy-methyl-ricinoleate in any of the chromatograms.

Table 16. Relative content (%) of the abundant fatty acids in transgenic tobacco shoots (S), calli generated on leaf explants (C), and calli obtained by NT-1 cell transformation (NT). Fatty acid methyl ester abbreviations as for Table 15.

			Fatty Acid Methyl Ester				
			16:0	18:0	18:1	18:2	18:3
constr.	tissue	identity					
B6	S	18-1	26.1	8.8	5.9	9.1	35.5
B6	S	10-3	25.7	9.2	3.7	8.1	39.0
B6	S	6-1	27.3	7.3	3.7	8.9	40.3
B6	S(c)	6-1	34.2	4.5	1.8	23.4	26.7
A4	S(c)	9-3	27.6	4.3	2.1	13.9	32.8
A4	S	9-3	25.2	9.4	7.0	6.7	33.4
A4	S	9-1	24.9	9.2	3.7	10.9	36.1
pBI121	S(c)	4/12-1	33.8	4.0	2.0	27.3	23.1
pBI121	S	4/12-1	25.3	7.9	5.6	7.8	36.8
pBI121	S	4/12-1	23.9	10.0	6.0	8.6	33.1
B6	C	10-4	31.2	5.1	2.9	34.9	18.3
B6	C	10-4	31.9	5.1	2.2	30.0	23.3
B6	C	10-5	30.1	7.1	6.0	28.6	19.5
A4	C	1-1	33.6	4.2	3.1	35.7	16.2
A4	C	9-4	30.6	5.0	3.6	36.5	15.0
A4	C	17-1	28.2	11.1	10.6	28.1	13.0
9/18 3	C	19-1	27.1	7.1	8.4	31.1	12.8
pBI121	C	4/12-2	33.0	4.8	3.3	34.0	18.9
B6	NT	1	29.0	9.7	16.7	31.4	8.5
B6	NT	2	22.4	8.1	8.6	15.0	10.1
A4	NT	3	25.8	8.6	17.7	25.4	9.4
A4	NT	4	23.7	7.9	11.1	18.7	10.9
pBI121	NT	5	28.8	7.8	15.3	37.1	7.5
pBI121	NT	6	36.8	5.8	8.2	36.1	7.7
pBI121	NT	7	34.2	6.4	7.6	36.5	6.2
pBI121	NT	8	34.4	6.1	8.5	32.8	6.2

CONCLUSIONS

As with expression of pFL2 in yeast, attempts to express pFL2 in plants have not yet yielded information supporting or contradicting the possibility that pFL2 encodes oleate-12-hydroxylase. So far, these transgenic plants have only been examined for the accumulation of ricinoleate, or other changes in the equilibrium fatty acid composition. A next experiment would be to determine whether pFL2 is in fact being expressed at the RNA level, by northern blot analysis. It would also be a high priority to assay the transgenic tissues for oleate-12-hydroxylase activity *in vitro*. The sensitivity of the radioisotopic *in vitro* assay may detect hydroxylase activity not detected by gas chromatography. Furthermore, expression of an oleate-12-hydroxylase in transgenic plant tissue may result in targetting of the ricinoleate produced to a degradation pathway. Plant microsomes contain a phospholipase with a preference for oxygenated acyl groups such as ricinoleic acid.[20] This may be part of a pathway for turnover of oxygenated fatty acids (such as those associated with membrane damage), and it is possible that the *in vitro* assay would separate hydroxylation from degradation, allowing ricinoleic acid to be detected.

REFERENCES

1. Coppens, N., Biosynthesis of fatty acids by seeds of *Ricinus communis*, *Nature*, 177, 279, 1956.

2. Canvin, D.T., Formation of oil in the seed of Ricinus communis L., *Can. J. Biochem. Physiol.*, 41, 1879-1885, 1963.

3. Yamada, M., and Stumpf, P.K., Enzymic synthesis of ricinoleic acid by extracts of developing Ricinus communis L. seeds, *Biochem. Biophys. Res. Commun.*, 14, 165-171, 1964.

4. James, A.T. Hadaway, H.C., and Webb, J.P.W., The biosynthesis of ricinoleic acid, *Biochem. J.*, 95, 448-452, 1965.

5. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

6. Panaccione, D.M., and Hanau, R.M., Characterization of two divergent β-tubulin genes from *Colletotrichum graminicola, Gene,* 86, 163-170, 1990.

7. Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W., GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants, *EMBO J.*, 6, 3901-3907, 1987.

8. Petit, A., Berkaloff, A., and Tempe, J., Multiple transformation of plant cells by *Agrobacterium* may be responsible for the complex organisation of T-DNA in crown-gall and hairy root, *Mol. Gen. Genet.*, 202, 388-393, 1986.

9. Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Perez-Grau, L., Kinney, A.J., Hitz, W.D., Booth, R., Schweiger, B., Stecca, K.L., Allen, S.M., Blackwell, M., Reiter, R.S., Carlson, T.J., Russell, S.H., Feldmann, K.A., Pierce, J., and Browse, J., Cloning of higher plant ω3 fatty acid desaturases, *Plant Physiol.*, 103, 467-476, 1993.

10. Newman, T.C., Ohme-Takagi, M., Taylor, C.B., and Green, P.J., DST sequences, highly conserved among plant *SAUR* genes, target reporter transcripts for rapid decay in tobacco, *Plant Cell,* 5, 701-714, 1993.

11. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D., *Molecular Biology of the Cell,* 2nd ed., Garland, New York, 1989, pp 414-415.

12. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M., and Somerville, C.R., Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*, *Science,* 258, 1353-1355, 1992.

13. Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S., and Somerville, C., A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana, J. Biol. Chem.*, in press.

14. Wada, H., Gombos, Z., and Murata, N., Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation, *Nature*, 347, 200-203, 1990.

15. Greenwood, J.S., and Bewley, J.D., Seed development in *Ricinus communis* (castor bean). I. Descriptive morphology, *Can. J. Bot.*, 60, 1751-1760, 1982.

16. Moore, L., Warren, G., and Strobel, G., Involvement of a plasmid in the hairy root disease of plants caused by *Agrobacterium rhizogenes*, *Plasmid*, 2, 617-626, 1979.

17. Koncz, C., and Schell, J., The promoter of T_L-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector, *Mol. Gen. Genet.*, 204, 383-396, 1986.

18. Ooms, G., Hooykaas, P.J.J., van Veen, R.J.M., van Beelen, P., Regensburg-Tuink, T.J.G., and Schilperoort, R.A., Octopine Ti-plasmid deletion mutants of *Agrobacterium tumefaciens* with emphasis on the right side of the T-region, *Plasmid*, 7, 15-29, 1982.

19. Cahoon, E.B., Shanklin, J., and Ohlrogge, J.B., Expression of coriander desaturase results in petroselinic acid production in transgenic tobacco, *Proc. Natl. Acad. Sci. USA*, 89, 11184-11188, 1990.

20. Banas, A., Johansson, I., and Stymne, S., Plant microsomal phospholipases exhibit preference for phosphatidylcholine with oxygenated acyl groups, *Plant Sci.*, 84, 137-144, 1992.

CHAPTER 7

CONCLUSIONS AND PERSPECTIVES

Different species of flowering plants synthesize varying amounts of triacylglycerol in their developing seeds. The majority of these species, including major oilseed crops, synthesize triacylglycerol in which the component fatty acids are similar to those found in other plant tissues, such as oleic, linoleic and linolenic acids. However, a number of plant species have been identified, as reviewed in chapter 1, which accumulate an unusual fatty acid in the seed triacylglycerols. Little is known about any function for such unusual fatty acids, and it is not a structural one, since these unusual fatty acids are generally excluded from membrane lipids. As in other oilseeds, one of the major metabolic activities of the developing endosperm of castor (*Ricinus communis*) seeds is the synthesis from imported sucrose of triacylglycerol. Approximately 90% of the component fatty acids of this storage lipid are the hydroxylated fatty acid, ricinoleic acid. This fatty acid is derived from oleic acid by hydroxylation at carbon 12. Ricinoleic acid is not produced by plants more commonly used in biological research, such as the model organism *Arabidopsis thaliana*, hence its biosynthesis must be studied in a plant such as castor. All previous studies of ricinoleic acid biosynthesis in plants have employed castor. Not surprisingly (given the amount of ricinoleic acid produced), the hydroxylase responsible for the transformation of oleic acid to ricinoleic acid is readily assayed in castor endosperm extracts. However, the nature of this hydroxylase was poorly

understood at the biochemical or molecular level. Characterization of the hydroxylase, and particularly, isolation of a gene (or genes) encoding the hydroxylase enzyme were the goal of this work.

Assays of oleate-12-hydroxylase were used in chapter 2 of this study to gain some new insight into the enzyme. The pattern of induction of the enzyme during seed ontogeny was investigated, and it was found that hydroxylase activity is detectable from the time of differentiation of the cellular endosperm tissue in which lipid is stored. It was conclusively shown for the first time that oxygen is a necessary cofactor for hydroxylation, and an approach employing stable isotopes and mass spectrometry was used to demonstrate that molecular oxygen is the immediate source of the hydroxyl oxygen atom at C12 of ricinoleic acid (Appendix A). Furthermore, it could be shown that antibodies against cytochrome b_5 completely inhibit oleate hydroxylation by castor endosperm extracts, indicating that cytochrome b_5 is the electron donor to the hydroxylase. The involvement of molecular oxygen and cytochrome b_5, and other properties of the hydroxylase reported in the literature, such as the identification of oleate esterified at *sn-2* of phosphatidylcholine as the immediate substrate,[1] are similar to the biochemical properties of the microsomal fatty acid desaturases.

Despite being readily assayed in crude endosperm extracts, oleate-12-hydroxylase was found to be labile during various attempts at fractionation of these extracts. This lability thwarted attempts at partial purification of the enzyme, and prompted the exploration of a number of alternative approaches to isolation of an

oleate-12-hydroxylase gene. These included attempts to obtain polyclonal antibodies against castor proteins which would inhibit hydroxylase activity (chapter 2), and screening yeast cultures expressing random castor endosperm cDNAs for ricinoleic acid production (chapter 3). Neither of these approaches was successful.

The results of some of the experiments described above, and a consideration of the possible iron-containing cofactors which could be involved in the hydroxylation mechanism, led to the development of a hypothesis concerning the molecular nature and evolutionary origin of the oleate-12-hydroxylase of castor. It appears that the ability to synthesize ricinoleic acid is found in isolated taxa of the plant kingdom, and there are no reports of the occurrence of ricinoleic acid in species closely related to castor. Therefore, it appears that the hydroxylase is a recently-evolved enzyme. Furthermore, the biochemical similarities betweeen the hydroxylase and the microsomal fatty acid desaturases of all plants suggests that the desaturases could be the progenitors of the hydroxylase. Recently, it has become known that the active site cofactor of at least one of the plant fatty acid desaturases is the diiron cluster previously characterised in methane monooxygenase[2]. This cofactor is likely to be found in all the plant fatty acid desaturases, and a discussion of the properties of this cofactor (chapter 4) reveals that, from a chemical viewpoint, it could also be the cofactor of a fatty acid hydroxylase, such as the oleate-12-hydroxylase of castor. Therefore, it is hypothesized that the castor oleate-12-hydroxylase is homologous to the plant microsomal fatty acid desaturases.

This hypothesis was used as the rationale for further experiments directed

toward isolation of an hydroxylase gene. Experiments described in chapter 4 involved use of one of the microsomal fatty acid desaturase genes, *fad3* of *Brassica napus*, in direct sceening of a castor developing-seed cDNA library. An apparently-conserved amino acid motif among widely-divergent fatty acid desaturases was also used to amplify sequences from castor endosperm mRNA using the polymerase chain reaction. In both these approaches, clones for what is concluded to be the castor *fad7* fatty acid desaturase were isolated, but no other desaturase-homologues (putative oleate-12-hydroxylase genes) were identified.

These and other results suggested that plant fatty acid desaturase genes were in some cases too divergent for such direct gene-isolation techniques. Still hypothesizing, however, that the hydroxylase is a desaturase-homologue, an experiment was designed that exploited recent advances in automated DNA sequencing technology. The partial sequencing of approximately 500 moderately abundant and seed-specific castor cDNAs allowed the identification of many genes expressed in castor endosperm. In particular, I was interested to identify clones with sequence similarity to desaturases, as candidates clones for oleate-12-hydroxylase. This is the first use of large-scale DNA sequencing as a strategy for the isolation of a particular gene. The use of this strategy was based on the expectation that oleate-12-hydroxylase transcripts should be moderately abundant in the developing endosperm, and that these could be further enriched by differential screening for seed-specific clones. Two of the clones sequenced, pCRS677 and pCRS834, had homology to plant membrane-bound fatty acid desaturases. In addition, a number of the other

genes identified in this experiment appear to encode proteins not previously isolated or characterized from higher plants, and will be a valuable resource for the work of numerous laboratories.

The clones pCRS677 and pCRS834 were characterized in experiments described in chapter 6, investigating the possibility that they encode oleate-12-hydroxylase. The two clones contained identical sequences and therefore appear to be derived from the same gene. The presence in the castor genome of a single copy of these sequences was confirmed by Southern analysis. The clone pCRS677 was used to isolate a class of abundant (1/560 to 1/1120) cDNAs, including the full-length cDNA clone pFL2, of which both strands were completely sequenced. This 1448 bp cDNA contains an 1161 bp open reading frame, encoding a 387 amino acid polypeptide with a calculated molecular weight of 44407. The nucleotide and deduced amino acid sequences have limited, but significant (ca. 47% and 38%, respectively) identity to known plant membrane-bound fatty acid desaturase genes. This indicated that pFL2 encodes either a desaturase-like protein, as oleate-12-hydroxylase is hypothesized to be, or a previously uncharacterized class of desaturase.

The abundance of clones isolated using pCRS677 as a probe was in agreement with the expression of pFL2 mRNA, as revealed by northern blot analysis. Transcripts hybridising to pFL2 are strongly expressed in developing endosperm, and very weakly expressed in leaf tissue, where ricinoleic acid is not synthesized, but where fatty acid desaturases are active. These data support the possibility that pFL2 encodes oleate-12-hydroxylase.

The clone pFL2, in a vector suitable for galactose-inducible expression in yeast, was transformed into yeast cells. These yeast cultures were analysed for fatty acid content, but no differences could be detected between those expressing the pFL2 cDNA and wild-type. However, control experiments demonstrating that pFL2 was being transcribed in the transformed yeast cells have not yet been done. The pFL2 cDNA was also inserted into a vector designed for expression in plants, and introduced into plant cells by three separate transformation procedures, all of which employed *Agrobacterium tumefaciens*. Carrot roots were cotransformed with a rooty-tumour inducing (Ri) plasmid, while tobacco leaf explants and tobacco cultured cells were transformed using *A. tumefaciens* strains harbouring disarmed Ti plasmids. An analysis of the fatty acid composition of the transgenic carrot root, tobacco shoot, and tobacco callus tissues obtained from these transformations failed to detect any ricinoleic acid, and no other changes in fatty acid composition (consistent, for instance, with pFL2-derived fatty acid desaturase activity) could be detected. Control experiments have not yet been done to confirm accumulation of RNA transcribed from pFL2 in the transgenic plant tissues.

It is concluded that the clone pFL2 has sequence and expression characteristics consistent with the possibility that it encodes oleate-12-hydroxylase, but this has not yet been shown. Despite having sequence similarity to fatty acid desaturases (and being strongly expressed in developing castor endosperm), pFL2 had no detectable effect on fatty acid content of transformed yeast or plants. Since ricinoleic acid is an unusually polar fatty acid, and not normally produced in vegetative tissues, or

incorporated into membrane-forming lipids, it is likely that production of large quantities of ricinoleic acid in tissues other than the developing castor endosperm could be detrimental to the cell. Poorly-understood mechanisms apparently operate in the castor endosperm to sequester ricinoleic acid in triacylglycerol, preventing its possible disruption of membrane structure. Vegetative plant tissues and growing yeast cells do not synthesize significant quantities of triacylglcerol, and may not have a mechanism for exclusion of unusual fatty acids, such as ricinoleic acid, from membrane-forming lipids. There are numerous possible reasons why ricinoleic acid was not detected in the yeast or plants transformed with pFL2, including the possibilities that pFL2 does not encode oleate-12-hydroxylase or that it was not expressed due to problems at the level of mRNA accumulation, translation or post-translational modification or targetting, or protein-protein interaction. However, supposing for a moment that expression of pFL2 resulted in the introduction of oleate-12-hydroxylase activity into the transgenic cells, then this was not associated with any obvious inhibition of growth, and it is an interesting possibility that these cells have mechanisms for recognition of the oxygenated fatty acid as alien, and for targetting this fatty acid for degradation, precluding its accumulation to detectable levels. A number of future experiments are suggested below which may help to differentiate between these possibilities.

Two obvious initial experiments are of the highest priority. These are to examine the transgenic tissues or cells for the accumulation of RNA transcribed from the pFL2 cDNA; and to assay the tissues or cells for ^{14}C-oleate-12-hydroxylase

activity using the *in vitro* assay described in chapter 2, or even simply by feeding ^{14}C-oleate *in vivo* for a brief period, after which the tissue could be killed and analysed for any labelled ricinoleic acid. A failure to accumulate RNA transcribed from the pFL2 cDNA would indicate that there was some problem at the level of transcription or transcript stability, and these areas would be the logical targets of further experiments designed to improve expression, such as the use of a different promoter, or a transcriptional enhancer.

Alternatively, detection of pFL2 transcripts, but no detectable enzyme activity would suggest two directions for investigation. It should be mentioned that the membrane-bound fatty acid desaturases of plants appear to be encoded by single genes, and the available genetic evidence in *Arabidopsis thaliana* gives no indication that more than one gene product is necessary for the desaturase component of the NADH reductase-cytochrome b_5-desturase system. However, none of the membrane-bound desaturases have been expressed in heterologous systems, and the possibility cannot be ruled out that expression of a single castor cDNA in another plant or other organism is sufficient for formation of a functional hydroxylase enzyme. This must be kept in mind, and re-evaluated as experimental characterization of the desaturases and of pFL2 proceeds.

In the first direction of investigation, expression of pFL2 could be attempted in *E. coli*, with the objective of purification of the encoded protein and preparation of polyclonal antibodies against this protein. Such antibodies could be used to determine whether pFL2 translation products accumulate in the transgenic tissues, and to

confirm that they accumulate only in the developing seeds of castor, and not in other parts of the plant. The developmental induction of such seed-specific expression in castor might be compared to the developmental induction of oleate-12-hydroxylase activity characterised in chapter 2. Furthermore, the ability of these antibodies to inhibit oleate-hydroxylation assays could be tested.

Relevant to the second direction of investigation, preliminary information not reported above, reveals that pFL2 is more closely related to the fad2 desaturase than other desaturases. This is the expected result if pFL2 encodes oleate-12-hydroxylase, since *fad2* encodes the microsomal Δ12 desaturase, biochemically most similar to the hydroxylase. However, it is also possible that pFL2 encodes a castor fad2 desaturase, not expressed, for some reason, in transgenic yeast, tobacco, or carrot. It is possible however that Δ12 desaturase activity could be detected in a tissue with a very low background level of this activity, and pFL2 might therefore be tested for complementation of the *Arabidopsis thaliana fad2* mutant.

If the experiments suggested earlier do detect oleate-12-hydroxylase activity in tissues expressing pFL2, and it can be established that this clone does indeed encode the hydroxylase, then an assortment of interesting experiments await the investigator. One would be to isolate the castor *fad2* gene, the presumed progenitor of the hydroxylase gene. As structure-function relationships of the fatty acid desaturases become better characterised, a comparison of the primary and predicted secondary structures of a desaturase and its derived hydroxylase homologue would be extremely interesting, and might provide valuable insight into the nature of the reaction

mechanisms of these enzymes, and suggest strategies for protein engineering.

Other interesting experiments would obviously be to work towards the envisioned future goal of this work, the production of ricinoleic acid in a plant not previously synthesizing this compound. Since it appears that one of the hurdles to be crossed may be to overcome endogenous mechanisms for degradation of alien fatty acids, this would be an interesting area for study. One experiment that might be suggested would be to transform *Arabidopsis* with the hydroxylase gene, and then to screen for mutants with a defect in the process of recognition or degradation of the alien ricinoleic acid. Another experiment might exploit an interesting technique which has not yet been used to its potential for the study of unusual fatty acids. This is the technique of Terzaghi,[3] in which exogenous fatty acids, as their Tween (polyoxyethelyenesorbitan) esters, painted on or otherwise applied to plant tissues, are extensively incorporated into membrane lipids. What would be the fate of ricinoleic acid applied in this way?

In any case, whether it is found that pFL2 encodes oleate-12-hydroxylase, or possibly some new class of desaturase, it seems that this clone will provide useful opportunities in the study of plant fatty acid metabolism. Furthermore, the many other clones, some with homology to genes previously uncharacterised in plants, identified in the course of isolating pFL2, should provide plant biologists with a number of other new opportunities.

REFERENCES

1. Bafor, M., Smith, M.A., Jonsson, L., Stobart, K., and Stymne, S., Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (Ricinus communis) endosperm, *Biochem. J.*, 280, 507, 1991.

2. Fox B.G., Shanklin, J., Somerville, C., and Munck, E., Stearoyl-acyl carrier protein Δ^9 desaturase from *Ricinus communis* is a diiron-oxo protein, *Proc. Natl. Acad. Sci.*, 90, 2486-2490, 1993.

3. Terzaghi, W.B., Manipulating membrane fatty acid compositions of whole plants with tween-fatty acid esters, *Plant Physiol.*, 91, 203, 1989.

APPENDIX

APPENDIX

MASS SPECTROMETRIC STUDIES: POSITION OF HYDROXYLATION AND SOURCE OF HYDROXYL OXYGEN

Note: the work presented in this appendix is largely that of Dr E.W. Underhill, my contributions being relatively minor. We acknowledge the assistance of Drs Tim Heath and Doug Gage with mass spectral analyses.

INTRODUCTION

Ricinoleic acid is formed by two pathways; by hydration of linoleic acid as is the case in the fungus *Claviceps purpurea* and by the hydroxylation of oleic acid as occurs in castor (see chapter 2). Yamada and Stumpf[1] reported ricinoleic acid was formed from ^{14}C-oleoyl-CoA by an NADPH-requiring soluble fraction (105 000 x g supernatant) isolated from extracts of the developing castor bean. However, in a following publication[2] it was claimed that the hydroxy product isolated above was not ricinoleic acid and, without reporting experimental data, indicated the product "has now been shown to be β-hydroxyoleic acid". Galliard and Stumpf[2] found that ricinoleic acid was derived from oleoyl-CoA by hydroxylation, that the oxidation required NADH and that the enzyme was present in the microsomal fraction of developing seeds of castor. In both papers, the methods used to identify the product

of the enzyme reaction were rigorous and appeared convincing.

Several methods are available which may be used to determine if ricinoleic acid is the product formed in a cell-free reaction mixture. Most frequently, chemical degradation followed by thin-layer and/or gas chromatographic separation of the resulting degradation products has been used for the identification of hydroxy fatty acids, including ricinoleic acid.[3-6] In the studies of Galliard and Stumpf[2] a very elegant system was employed in which both ^{14}C- and ^{3}H-labeled oleoyl-CoA were utilized. A similar method was considered for the studies reported here and an attempt was made to obtain ^{3}H-labeled oleic acid from the same source as had been used by Galliard and Stumpf. However, the material was no longer available and its synthesis is not a trivial matter.

An alternative and unequivocal method of identifying the hydroxylated product formed in these studies was investigated and it was decided to employ mass spectrometry. The use of mass spectrometry for reaction product analysis is an attractive method for a number of reasons. These include: (1) The electron impact mass spectra (EI/MS) of the trimethylsilyloxy (TMS) derivatives of several hydroxylated fatty acid methyl esters, including ricinoleic acid, have been reported.[7] It has been shown that the position of the hydroxyl group within the chain may be readily ascertained from the fragmentation ions produced. (2) Mass spectrometry, particularly when combined with the high resolution capability of capillary GC, has high sensitivity and can be used for reaction product identification as opposed to product confirmation as is often the case with chemical degradation studies. (3)

Chemical degradation is, at best, a confirmatory procedure that offers little if the product of the enzyme reaction is other than that expected. In addition, chemical degradation is often followed by thin-layer and/or packed column separation of the reaction products, procedures that have relatively low resolving power by current standards. (4) In addition to providing structural identification of the product formed, mass spectrometry in combination with the use of stable isotopes may also be used to determine the source of oxygen in the hydroxyl group of ricinoleic acid.

EXPERIMENTAL DESIGN

The EI mass spectrum of the TMS-derivative of methyl-ricinoleate (mol. wt. 384) contains a number of fragmentation ions which characterize the compound. These include the ions at m/z 369 $[M-15]^+$, 353 $[M-(CH_3O)]^+$, 187 (base peak) and 299 (the last 2 are cleavage ions alpha to the TMS group) and the rearrangement ion at 270 $[CH_3O(OTMS)C(CH_2)_7CH=CH-CH_2]^+$. The ions at m/z 187, 299 and 270 unequivocally establish the hydroxyl at C-12. Accordingly, ricinoleic acid or any other hydroxylated product formed from a specifically deutero-labeled oleoyl-CoA, would be recognisable in a reaction mixture from the resulting fragmentation ions present in the EI mass spectrum of the TMS-methyl ester(s). In addition, since the hydroxyl oxygen in the product is not lost as a result of TMS-derivatization, it should be possible to determine if the hydroxyl oxygen is derived from molecular oxygenby conducting the experiment in the presence of oxygen-18. Since the ion at m/z 187

constitutes the base peak in the spectrum a search was made for a source of oleic acid containing deuterium at carbons 12 through 18. Thus, for ricinoleic acid formed from an oleoyl-CoA labeled with deuterium at carbons 12 through 18, the mass of the fragmentation ion corresponding to the ion at m/z 187 $[CH_3(CH_2)_5CHOTMS]^+$ would be increased by (1) the number of deuteriums in the substrate and by (2) two additional mass units if the oxygen was derived from $^{18}O_2$.

MATERIALS AND METHODS

Chemicals

1-^{14}C-oleoyl-CoA (52 μCi/μmol) was prepared using acyl-CoA synthetase (Sigma) and 1-^{14}C-oleic acid (1 μmol) according to the method of Taylor *et al.*[8] The product was formed in >95% yield and had a radiopurity of 98%. Trideuterooleoyl-CoA was synthesized in a similar manner starting from 18-trideuterooleic acid (1 μmol, >98% isotope purity): the deutero-fatty acid, synthesized by Dr. A.P. Tulloch (deceased), was provided by the Plant Biotechnology Institute, Saskatoon, SK. Canada. All other chemicals were obtained from commercial sources.

Enzyme Preparation

Developing castor endosperm plus embryo tissue at stages IV to VI[9] was

collected into liquid nitrogen. This material was powdered with a mortar and pestle, and stored at -70°C. This powdered endosperm (1.41 g) was extracted by grinding (Ten Broek homogenizer) with 1.4 ml of extraction buffer consisting of 0.05 M PIPES pH 7.1, 10% (w/v) glycerol, leupeptin (5 μg/ml) and 1 mM cysteine. The residue was extracted a second time with a similar volume of extraction buffer. The enzyme extracts were combined and centrifuged at 13 600 x g for 15 min in a microfuge. The clear supernatant ($\sim$ 3 ml) was decanted leaving behind the floating pad of fat and residual solids. $MgCl_2$ was added to give a final concentration of 50 mM and the mixture centrifuged at 13 600 x g for 10 min. The supernatant was used as the source of enzyme. The enzyme was used immediately, or was dropped into liquid nitrogen and stored at -70°C.

Enzyme Assays

Reaction mixtures contained 50 μl enzyme, 10 μl 20 mM NADH in Tris-HCl pH 7.1, 5 μl of substrate [1-^{14}C-oleoyl-CoA (15 950 dpm/μl; 52 Ci/mol) or 18-trideuterooleoyl-CoA (0.2 nmol/μl)] and 435 μl of 50 mM PIPES pH 7.1 containing 10% (w/v) glycerol. For reactions carried out in the presence of $^{18}O_2$, substrate, cofactor and buffer were first added, and immediately following the addition of enzyme, the reaction tubes were alternately evacuated and filled with one of the following: (1) a mixture of 20% $^{18}O_2$ plus 80% nitrogen, (2) nitrogen, or (3) air. Replicated reactions were carried out at 30°C and were terminated by addition of 0.5

ml of 15% KOH in methanol. Mixtures were saponified at 80°C for 30 min, cooled, acidified with 0.5 ml 2.5 M HCl and the liberated acids extracted using 3.5 ml hexane:ispropanol (3:2) and 2.5 ml 0.2 M Na_2SO_4[10]. The organic phase was separated and concentrated under nitrogen.

Quantitation of Ricinoleic Acid

Ricinoleic acid was determined by GC using heptadecanoic acid as an internal standard; for routine analyses the C-17 internal standard was added to the reaction mixtures immediately prior to saponification. Methyl esters of fatty acids were prepared after reducing the organic phase to dryness by the addition of 0.5 ml of anhydrous 1 M HCl in methanol and heating to 80°C for 1 h. Methyl esters were recovered from the cooled reaction mixtures by addition of 1.5 ml 0.9% NaCl and extraction into 200 μl hexane.

TMS-derivatives of hydroxy fatty acids were prepared by adding an excess (~ 5 μl) of trimethylsilylimidazole (Pierce) to a dried (Na_2SO_4) solution of methyl esters in hexane. Derivatization was carried out at room temperature for a period of at least 2 h. Excess derivatizing reagent was removed under a stream of dry nitrogen and the TMS-methyl esters dissolved in hexane.

Thin Layer Chromatography

Fatty acids were spotted (alongside authentic oleic acid and ricinoleic acid standards) on a silica TLC plate (Baker Si250). The plate was developed in a paper-lined tank containing benzene:ethyl ether:ethanol (100:30:2).[11] Fatty acids were detected by staining with iodine vapour. Radioactivity could be quantitated by autoradiography followed by scraping and scintillation-counting of the oleic acid and ricinoleic acid spots, but equivalent quantitation was routinely obtained using a BioScan 2000 scanner.

Mass Spectrometry

GC-MS of TMS derivatives of fatty acid methyl esters was performed on a Finnigan TSQ 70 mass spectrometer equipped with a Varian 3400 gas chromatograph. The column used was a DB-% capillary (30 m x 0.25 mm: J & W Scientific Inc., Ranchero Cordova, CA), programmed from 150 to 280°C at 12°C min^{-1}. Helium was used as carrier gas (~ 1 ml min^{-1}). Spectra were aquired every 0.5 s over the range m/z 50 to 400.

Mass spectral analyses were carried out on duplicate samples of enzyme incubations containing 18-trideuterooleoyl-CoA and each sample was assayed in triplicate by GC-MS. The results were averaged and standard deviations calculated.

RESULTS AND DISCUSSION

The crude enzyme preparation used contained appreciable quantities of ricinoleic acid; the amount of ricinoleic acid (calculated as methyl ricinoleate) was 5.88 ± 0.56 $\mu g(50\ ul)^{-1}$ (n=6).

The enzyme preparation was active. Bioscan data, obtained from three incubations using 1-^{14}C-oleoyl-CoA as substrate, showed 20.15% (20.66, 20.60 and 19.19) of the activity supplied had been converted into a component co-migrating with ricinoleic acid. Based on the total activity and the specific activity of 1-^{14}C-oleoyl-CoA employed, the amount of labelled product (ricinoleic acid) formed in these calculations was calculated to be 0.14 nmol or 41.5 ng. (Note: labelled ricinoleate constitutes ~ 0.7% of the total ricinoleate in the sample).

Mass spectral data obtained from incubations containing 18-trideuterooleoyl-CoA unequivocally demonstrated that the substrate was converted to ricinoleic acid. As indicated above, alpha-fragmentation to the TMS group at the 12-position of ricinoleic acid gives rise to the ion at m/z 187, the base peak in the mass spectrum. The corresponding fragmentation ion which could be derived from ricinoleic acid formed from the trideuterated substrate would occur at m/z 190. From the % relative abundance of ions in the range m/z 187 to 192 (Table 17) the ion at m/z 190 is significantly greater than that found in "blank" incubations (1.356 vs 0.631). From this data the amount of the deutero-labeled substrate converted can be estimated, namely (1.356 - 0.631)(5880)/100 = 42.6 ng, in close agreement with the calculated

conversion of the ^{14}C-labeled substrate.

The aquired mass-spectral data was searched using single ion monitoring to determine if another mono-trimethylsilyloxy, monounsaturated C-18 methyl ester could be detected. Those searched included ions at m/z 384 [M^+], 369 [M^+], as well as ions corresponding to alpha cleavages of 2- and 3-trimethylsilyloxy-methyl oleate, with and without a trideuterated carbon. No evidence was obtained in support of these components. In the course of these searches, data was obtained supporting the presence of 2-hydroxystearic; co-incident retention time and matching spectrum with authentic 2-trimethylsilyloxy-methy stearate.

Data obtained from experiments employing 1-^{14}C- and 18-D_3-oleoyl-CoA confirmed and extended findings reported previously.[2] Labeled carbon from 1-^{14}C-oleoyl-CoA was extensively incorporated into ricinoleic acid (or a component co-migrating with ricinoleic acid) when reactions were carried out in air or in a mixture of 20% $^{18}O_2$ plus 80% nitrogen. In air, 16.6% (18.7, 15.4, 15.7) of ^{14}C co-migrated with authentic ricinoleic acid. In 20% $^{18}O_2$, 15.0% (15.4, 15.1, 14.6) of ^{14}C co-migrated with authentic ricinoleic acid. There was no conversion of ^{14}C into the product in the absence of oxygen.

The results obtained from reaction mixtures containing 18-trideuterooleoyl-CoA, incubated in the presence of $^{18}O_2$, conclusively demonstrated that the hydroxyl oxygen of ricinoleic acid was derived from the $^{18}O_2$. Data obtained from these incubations (Table 17) show a significant increase in the relative abundance of the ion at m/z 192, as would be expected from a trimethylsilyloxy alpha fragmentation ion

containing an ^{18}O and 3 deuterium atoms. The slightly elevated relative abundance of the ion at m/z 189 is likely a result of ^{18}O incorporation into a ricinoleic acid intermediate present in the enzyme preparation.

Table 17. Formation of ricinoleic acid from oleoyl-CoA by cell-free extracts of developing castor endosperm. Extracts were incubated with 18-trideutero-oleoyl-CoA in the presence of air or 18O2. The position at which oleic acid was hydroxylated as well as the source of the hydroxyl oxygen was determined by GC-MS analysis of fragmentation ions (m/z 187 to 192) derived from the recovered trimethylsilyl ether of methyl ricinoleate. Each treatment was replicated two times and each replicate was analyzed three times by GC-MS.

Experiment	% Relative Abundance of Ions[a] at m/z					
	187	188	189	190	191	192
1-^{14}C-oleoyl-CoA	100	14.859 (0.227)[b]	4.722 (0.144)	0.631 (0.076)	0.075 (0.062)	0.011 (0.014)
D_3-oleoyl-CoA (air)	100	15.302 (0.405)	4.773 (0.086)	1.356 (0.059)	0.183 (0.013)	0.050 (0.024)
D_3-oleoyl-CoA ($^{18}O_2$)	100	15.094 (0.263)	5.687 (0.082)	0.880 (0.122)	0.130 (0.061)	0.525 (0.069)

[a] mean values, n=6

[b] standard error of the mean

REFERENCES

1. Yamada, M., and Stumpf, P.K., Enzymic synthesis of ricinoleic acid by extracts of developing *Ricinus communis* L seeds, *Biochem. Biophys. Res. Commun.*, 14, 165-171, 1964.

2. Galliard, T., and Stumpf, P.K., Fat metabolism in higher plants XXX. Enzymatic synthesis of ricinoleic acid by a microsomal preparation from developing Ricinus communis seeds, *J. Biol. Chem.*, 241, 5806, 1966.

3. Barucha, K.E., and Gunstone, F.D., Vegetable oils VI: The component acids of ergot oil, *J. Chem. Soc.*, 610-614, 1957.

4. Ahmad, M.S., Ahmad, M.U., and Osman, S.M., A new hydroxyolefinic acid from *Plantago major* seed oil, *Phytochem.*, 19, 2137-2139, 1980.

5. Husain, S., Ahmad, M.U., and Osman, S.M., New hydroxy fatty acid from seed oil of *Baliospermum axillare*, *Phytochem.*, 19, 75-77, 1980.

6. Ahmad, M.U., Husain, S.K., and Osman, S.M., Ricinoleic acid in Phyllanthus niruri seed oil, *J. Am. Oil Chem. Soc.*, 58, 673, 1981.

7. Kleiman, R., and Spencer, G.F., Gas chromatography-mass spectrometry of methyl esters of unsaturated oxygenated fatty acids, *J. Am. Oil Chem. Soc.*, 50, 31-38, 1973.

8. Taylor, D.C., Weber, N., Hogge, L.R., and Underhill, E.W., A simple enzymatic method for the preparation of radiolabeled erucoyl-CoA and other long-chain fatty acyl-CoAs and their characterization by mass spectrometry, *Anal. Biochem.*, 184, 311-316, 1990.

9. Greenwood, J.S., and Bewley, J.D., Seed development in *Ricinus communis* (castor bean). I. Descriptive morphology, *Can. J. Bot.*, 60, 1751-1760, 1982.

10. Moreau, R.A., and Stumpf, P.K., Recent studies of the enzymic synthesis of ricinoleic acid by developing castor beans, *Plant Physiol.*, 67, 672, 1981.

11. Donaldson, R.P., Accumulation of free ricinoleic acid in germinating castor bean endosperm, *Plant Physiol.*, 59, 1064-1066, 1977.