Isolation and Characterization of Two Safflower Oleoyl-Acyl Carrier Protein Thioesterase cDNA Clones

Deborah S. Knutzon¹, Janice L. Bleibaum, Janet Nelsen, Jean C. Kridl*, and Gregory A. Thompson

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

ABSTRACT

Oleoyl-acyl carrier protein (18:1-ACP) thioesterase has been partially purified from developing safflower (Carthamus tinctorius) seeds. Protein species with molecular masses of 34 and 40 kD associated with thioesterase activity were identified and partially sequenced. Analysis of amino-terminal and internal cyanogen bromide peptide sequences revealed no differences in the primary structure of the two species. Amino acid sequence was used to design degenerate oligonucleotides for primers in a polymerase chain reaction (PCR) using safflower embryo cDNA as a template. A 380-base pair PCR product was used to isolate two classes of cDNA clones, designated 2-1 and 5-2, from the embryo cDNA library. Clone 2-1 encodes a 389-amino acid protein including a 60-amino acid transit peptide, and contains all of the protein sequence determined from the 34- and 40-kD proteins. Clone 5-2 encodes a 385-amino acid protein with 80% identity to that encoded by 2-1. Expression of the two safflower cDNA clones in Escherichia coli resulted in a 50- to 100-fold increase in the level of 18:1-ACP thioesterase activity. Both thioesterases are most active on 18:1-ACP: however, the enzyme encoded by 5-2 shows less discrimination against saturated 16- and 18-carbon acyl-ACP substrates.

Fatty acids are important components of membrane phospholipids in all plant tissues and are also stored in the form of triacylglycerols in the seeds of many species for use as an energy source upon germination. Fatty acid biosynthesis in plants is catalyzed by a series of soluble, discrete enzymes (23) that are localized in the chloroplasts of leaves (16) and the proplastids of developing seed tissue (14, 25, 31). Fatty acids up to 18 carbon atoms in length are synthesized in twocarbon increments as ACP² thioesters (15). Acyl-ACP thioesterase (EC 3.1.2.14) catalyzes the hydrolysis of these acyl-ACP thioesters (24) to release free fatty acids, which are transported to the cytoplasm for further modifications and incorporation into storage triacylglycerols (28). Whereas most plants accumulate predominantly 16- and 18-carbon fatty acids in their seeds, some species are notable for the accumulation of shorter-chain fatty acids. An example is California bay (*Umbellularia californica*), which accumulates high levels of lauric acid (12:0) in its seed triacylglycerols. Recently, a 12:0-ACP-specific thioesterase has been isolated from this species and shown to redirect fatty acid synthesis to the production of laurate in transgenic plants (6, 18, 30). In plants that accumulate the longer-chain fatty acids, the specificity of the acyl-ACP thioesterase may also play a role in determining the final fatty acid composition of the storage oil.

We are interested in studying acyl-ACP thioesterases with specificities for 16- and 18-carbon fatty acids. Acyl-ACP thioesterase activities have been characterized at various stages of purification from a number of plant species that accumulate 16- and 18-carbon fatty acids in their triacylglycerols, including avocado (17, 24), safflower (*Carthamus tinctorius*) (12), squash (8), and rapeseed (7). The enzymes show distinct preferences for 18:1-ACP over 18:0-ACP and 16:0-ACP substrates and are relatively inactive on acyl-CoAs.

To investigate further the role of acyl-ACP thioesterase in the determination of seed oil composition, we have purified and obtained partial amino acid sequence of an 18:1-ACP thioesterase from developing safflower seeds. Two different classes of cDNA clones encoding 18:1-ACP thioesterase have been isolated and characterized. The substrate specificity of the protein encoded by each clone has been examined by expression in *Escherichia coli*.

MATERIALS AND METHODS

Materials

 $[1-^{14}C]$ Acyl-ACP substrates were synthesized by the enzymic method described by Rock et al. (21) using ACP that was purified from *E. coli* K12 by the method of Rock and Cronan (20). Radioactive fatty acids (54–58 mCi/mmol) were purchased from Amersham. Immature safflower (*Carthamus tinctorius*) seeds were collected from greenhouse-grown plants 16 to 18 d postanthesis, frozen in liquid nitrogen, and stored at -70°C. Oligonucleotide primers for PCR and in vitro mutagenesis were synthesized on an Applied Biosystems model 380A DNA synthesizer.

Purification and Sequencing of 18:1-ACP Thioesterase

A partial purification of 18:1-ACP thioesterase from immature safflower seeds by acetone powder extraction, acid precipitation, and ACP-Sepharose chromatography was carried out as described by McKeon and Stumpf (12). Fractions from the ACP-Sepharose column that contained acyl-ACP thioesterase activity were pooled and subjected to further

¹ Present address: Department of Molecular Biology, Genentech, Inc., 460 Point San Bruno Blvd., South San Francisco, CA 94080.

² Abbreviations: ACP, acyl carrier protein; CHAPS, 3-{(3-cholamidopropyl)dimethylammonio}-1-propanesulfonate; CNBr, cyanogen bromide; PCR, polymerase chain reaction; IPTG, isopropyl β -Dthiogalacotopyranoside; PVDF, polyvinylidenedifluoride; pI, isoelectric point.

purification by chromatofocusing. The buffer was changed to buffer A (20 mm bisTris-HCl, pH 7.4, 20% [v/v] glycerol, 0.1% [w/v] CHAPS) by ultrafiltration on a PM10 membrane in an Amicon pressure cell. The sample was then applied at 0.6 mL/min to a 1 cm (i.d.) × 7 cm column of Pharmacia PBE94 ion exchanger that had been equilibrated at 1.2 mL/ min in buffer A. The column was washed with 5 bed volumes of buffer A, and the proteins were eluted at 1 mL/min in a pH 7 to pH 4 gradient by the application of buffer B (10% [v/v] Polybuffer 74, 20% [v/v] glycerol, 0.1% [w/v] CHAPS, pH 4). Two-milliliter fractions were collected and assayed for thioesterase activity without pH adjustment.

Proteins eluted from the PBE94 chromatofocusing column were analyzed by SDS-PAGE on 8 to 16% Tris-glycine gels (NOVEX, San Diego, CA) and silver stained (3). Fractions 22 to 33 were combined to form pool A and fractions 34 to 50 were combined to form pool B. Proteins from the two pools were concentrated, separated by preparative SDS-PAGE, and electroblotted to PVDF membranes (Immobilon-P, Millipore) in a buffer consisting of 125 mM Tris base, 25 mM glycine, and 10% [v/v] methanol. For protein bands identified as thioesterase candidates, approximately 10% of the blotted protein from pool B was used for N-terminal sequence analysis. Sequencing was performed with an Applied Biosystems 477A protein sequencer with an on-line 120A PTH analyzer. The remaining blotted proteins from candidate bands in each pool were subjected to CNBr cleavage using a Probe-Design Separation System (Promega). CNBr peptides were separated using 10 to 20% Tris-Tricine gels (NOVEX), electroblotted to ProBlott (Applied Biosystems), and sequenced as described above. Each sequence is designated by a sequence number, e.g. SQ828.

Enzyme Assay

Thioesterase assays were performed during protein purification as described by Ohlrogge et al. (17) at pH 8.5 in 100 mM Tricine-NaOH, pH 8.5, using 3 μ M [¹⁴C]acyl-ACP substrate. To conserve the more expensive [1-¹⁴C]18:1-ACP, we used [1-¹⁴C]18:0-ACP as a substrate throughout the protein purification steps, and [1-¹⁴C]18:1-ACP was used to verify the substrate specificity of the enzyme.

Thioesterase assays used in *E. coli* expression experiments were performed as described by Davies et al. (6) in 7 mM NaH₂PO₄-NaOH, pH 8.0, 20% [v/v] glycerol, 1 mM DTT, and 0.02% [v/v] Triton X-100, with 0.5 μ M [1-¹⁴C]acyl-ACP substrate. Crude safflower extract, for use as a control, was prepared by homogenizing 250 mg of frozen, powdered safflower seeds in 1 mL of homogenization buffer consisting of 50 mM NaH₂PO₄-NaOH, pH 7.5, 2 mM DTT, 2 mM sodium ascorbate, 20% [v/v] glycerol, 1% [w/v] PVP-10, and 5 mM diethyldithiocarbamate. One unit of enzyme activity catalyzes the hydrolysis of 1 μ mol [¹⁴C]acyl-ACP/min at 30°C.

Library Construction and Screening

Construction of a plasmid cDNA bank from mRNA of developing safflower embryos, insertion of the cDNA clones into λ gt10, and phage screening were as described (29).

An aliquot of the plasmid cDNA bank was linearized with

EcoRI and used as template for PCR. The following degenerate oligonucleotides were derived from amino acid sequence and used as primers: 2683 (from SQ828), 5'-AAGG ATCCTAC/TAAA/GGAA/GGTA/C/G/TTTC/TATC/T/ AAT-3'; 2677 (from SQ829), 5'-CCAAAGCTTTA/GTTA/C/ G/TACA/GTGC/TTTA/GTTCAT-3'; 2676 (from SQ834), 5'-GTTGTCGACA/GTCA/GTCA/GTTA/C/G/TACC/T TTC/TTG-3'. PCR reactions (50 µL) contained 50 mM KCl, 10 mм Tris-HCl, pH 8.3, 1.5 mм MgCl₂, 0.01% [w/v] gelatin, 200 µM each deoxyribonucleotide triphosphates, 100 pmol of each primer, and 20 ng of template. The reaction mix was heated to 100°C for 3 min and 1.5 unit of Amplitaq DNA polymerase (Perkin Elmer Cetus) was added. Amplification took place in a Biosycler oven (BIOS Corporation, New Haven, CT) using the following thermal profile: six cycles of 2 min at 95°C; 1 s at 65°C; 2 min ramp to 55°C; 30 s at 55°C; 30 s at 72°C, followed by 37 cycles of 15 s at 95°C; 1 s at 65°C; 2 min ramp to 55°C; 15 s at 55°C; 15 s at 72°C.

The complete nucleotide sequence of both strands of the cDNA clones was determined by the method of Sanger et al. (22) using double-stranded plasmid DNA as template. Sequence analysis was performed using Intelligenetics (Mountain View, CA) and MacVector (IBI, New Haven, CT) computer software.

Expression of cDNA Clones in E. coli

In vitro mutagenesis of safflower 18:1-ACP thioesterase clones 2-1 and 5-2 was performed using the method of Kunkel et al. (10). The following oligonucleotides were designed to direct the insertion of Sall and Ncol sites immediately at the start of the mature protein coding region of cDNA clones 2-1 and 5-2, respectively: 3135, 5'-ACGGTGGCGCCGGTGTTGTCGACCATGGCGGTGA-GGACC-3'; 3134, 5'-GTTTCTCCGTTGCTAATGTCGA-CCATGGCGGTGGCGACCGGA-3'. The mature coding region plus 3'-untranslated sequences in 2-1 and 5-2 cDNA clones were removed as an NcoI/SmaI fragment and inserted into pET8c (27) that had been cut with BamHI, treated with Klenow fragment of DNA polymerase to create a blunt end, and then cut with Ncol. The resulting expression constructs, pCGN3270 (2-1) and pCGN3271 (5-2), were designed to express the mature safflower 18:1-ACP thioesterase cDNA sequences directly from the T7 promoter. For expression analysis, the constructs were transferred into E. coli strain BL21(DE3) containing the T7 RNA polymerase gene under control of the IPTG-inducible lacUV5 promoter (27).

For thioesterase activity assays, cells containing pCGN3270, pCGN3271, or pET8c as a control were grown at 37°C to an A_{600} of approximately 0.5 in 2YT medium (16 g of tryptone, 10 g of yeast extract, and 5 g of NaCl per L, pH 7.0) containing 0.4% [w/v] glucose and 300 μ g/mL of penicillin. Induction was achieved by the addition of IPTG to 0.4 mM and growth for a further 1.5 h. Ten-milliliter aliquots of culture were harvested, and pelleted cells were stored at -70° C. Prior to assay, pellets were resuspended in 500 μ L of thioesterase assay buffer and sonicated for three bursts of 20 s each. Protein concentrations were determined using the Bio-Rad Protein Assay.

RESULTS

Purification and Sequencing of 18:1-ACP Thioesterase

18:1-ACP thioesterase was partially purified from developing safflower seeds according to the method of McKeon and Stumpf (12). Material derived from the final ACP-Sepharose step of this protocol still contained a substantial number of proteins (Fig. 1B, lane T). To purify further the thioesterase, an additional chromatofocusing step was performed. Two peaks of 18:0-ACP thioesterase activity were eluted from the chromatofocusing column with the pH 7 to pH 4 gradient (Fig. 1A). Fractions 22 to 33 and fractions 34 to 50 were pooled (pools A and B, respectively), concentrated, and assayed to verify 18:1-ACP activity. In each of the protein pools, the 18:1-ACP thioesterase activity was 2.5-fold greater than the corresponding 18:0-ACP activity (data not shown). Analysis of chromatofocusing column fractions by SDS-PAGE revealed prominent proteins of approximately 34 and 40 kD in the fractions corresponding to both peaks of thioesterase activity (Fig. 1B).



Figure 1. Chromatofocusing of partially purified 18:1-ACP thioesterase from immature safflower seeds. A, Pooled fractions from an ACP-Sepharose column were applied to a PBE94 chromatofocusing column and eluted with a gradient of pH 7 to pH 4. The beginning and end of the pH gradient are indicated by arrows. Two-milliliter fractions were collected and assayed for 18:0-ACP thioesterase activity. Activity is reported as milliunits (mU)/fraction. One milliunit of enzyme catalyzes the release of 1 nmol [1-14C]18:0/min at 30°C. B, Proteins eluted from the chromatofocusing column were subjected to SDS-PAGE and visualized by silver staining. Lanes T and FT represent the total mixture applied to the PBE94 column and the flow-through, respectively. Numbers above lanes indicate the column fraction number. The positions of the 40- and 34-kD proteins excised from each pool are indicated by the arrows. Positions of molecular mass standards (in kD) are indicated to the right of the gel.



Figure 2. Partial amino acid sequencing of safflower 18:1-ACP thioesterase. The CNBr peptides (SQ828, SQ829, SQ833, SQ834) of the 34- and 40-kD safflower 18:1-ACP thioesterase proteins were aligned relative to the amino-terminal sequence, SQ830. The amino acid sequences (in one-letter code) used to design degenerate PCR primers 2676, 2677, and 2683 are underlined. Arrowheads indicate the direction of PCR extension resulting from each primer. The determined amino acid sequence of each peptide is shown in Figure 3. Position of the carboxy terminus of the 34-kD protein is inferred by the mol wt. Numbers above the line map indicate the approximate distances between nucleotides encoding the peptides expected in the cDNA.

Proteins from pools A and B were separated on preparative SDS-PAGE and blotted to PVDF membranes for aminoterminal sequencing and CNBr cleavage. Amino-terminal sequence analysis was performed on the uncleaved 34- and 40-kD proteins from pool B, and the sequences obtained were found to be identical. The 34- and 40-kD proteins from pools A and B were also subjected to CNBr cleavage, and the resulting peptides were partially sequenced. Sequences of the CNBr peptides from the 34-kD protein were identical to those obtained from the 40-kD protein, and no differences were seen in sequences arising from pool A versus pool B proteins. Based on incomplete cleavage of some of the peptides, their approximate mol wts, and partial homology to a 12:0-ACP thioesterase from California bay (30), the CNBr peptide sequences were aligned with respect to the amino-terminal sequence, SQ830 (Fig. 2).

Generation of an 18:1-ACP Thioesterase DNA Probe Based on Protein Sequence

The PCR method was used to generate a nucleic acid probe with which to isolate cDNA clones encoding 18:1-ACP thioesterases. Degenerate oligonucleotides were designed that contained all possible codons corresponding to the regions of the peptides indicated in Figure 2. Amplification of linearized cDNA template using primers 2683 and 2677 (based on peptides SQ828 and SQ829, respectively) resulted in an approximately 600-bp fragment, as visualized by agarose gel electrophoresis. This is consistent with the fragment size expected based on the peptide alignment. To confirm that this fragment contained 18:1-ACP thioesterase sequences, a second round of PCR was performed using the 600-bp fragment as template and an oligonucleotide derived from an internal peptide sequence (SQ834) as a primer. The product of this reaction was approximately 380 bp, a size again consistent with the peptide map. As an additional

confirmation, the 600-bp PCR product was subcloned and sequenced. The deduced amino acid sequence of this fragment contained peptide SQ834 as well as a minor peptide, SQ834B, not previously placed on the map (data not shown).

Isolation of cDNA Clones Encoding 18:1-ACP Thioesterase

The gel-purified 380-bp PCR product was radiolabeled and used to probe a safflower embryo cDNA library. Approximately 130,000 phage plaques were screened and 6 positive plaques were identified. After plaque purification, plasmid DNA was recovered from the six clones and analyzed by restriction mapping and partial DNA sequencing. Two classes of cDNA clones were identified based on restriction mapping. One cDNA clone, 5-2, contained recognition sites for several common restriction enzymes, whereas the other five cDNA clones contained none of these sites. Partial DNA sequence obtained from the 5' ends of the five similar clones indicated that they were different-length versions of each other. The longest clone of this class is identified as 2-1.

The complete nucleotide sequences of cDNA clones 2-1 and 5-2 were determined. The 1561-bp sequence of cDNA clone 2-1 is presented in Figure 3 along with its deduced amino acid sequence. Beginning with the putative translation initiation signal at nucleotide 173, which is similar to the plant consensus sequence (11), the open-reading frame extends 1167 bp, encoding a 389-amino acid preprotein. All of the sequenced peptides from the 34- and 40-kD isoforms are accounted for in this open-reading frame. Based on the location of the N-terminal peptide sequence, SQ830, the 2-1 cDNA encodes a transit peptide of 60 amino acids. This is consistent with the expected localization of 18:1-ACP thioesterase to the seed proplastid (31). The calculated mol wt for the mature protein is 37,254.

The complete nucleotide sequence of cDNA clone 5-2 was also determined (Fig. 3). This cDNA is somewhat shorter than 2-1 but is highly conserved within the protein coding region. Clone 5-2 contains 1449 bp and encodes an openreading frame of 1155 bp beginning with the first in-frame ATG at position 94. Assuming a transit peptide cleavage site analogous to 2-1, 5-2 also encodes a 60-amino acid transit peptide, but the mature protein has only 325 amino acids and a calculated mol wt of 36,981. The deduced amino acid sequence encoded by the 5-2 cDNA, although homologous to that of 2-1, does not contain sequences that exactly correspond to any of the determined peptide sequences and therefore does not encode any of the observed protein isoforms from chromatofocusing.

Comparison of the deduced amino acid sequences of the two 18:1-ACP thioesterase cDNA clones indicates that the mature proteins are 82% identical, and the corresponding DNA sequences share 80% identity. Despite this high degree of homology, computer estimates of the pI of the two proteins differ considerably. The estimated pI for the mature protein encoded by 2-1 is 5.8, whereas that of the protein encoded by 5-2 is 8.1.

Expression of 18:1-ACP Thioesterase in E. coli

The safflower thioesterase cDNAs were expressed in *E. coli* using an IPTG-inducible T7 polymerase system (27). The

segments of the 2-1 and 5-2 cDNAs encoding the mature form of the protein were inserted into pET8c (27) to create pCGN3270 and pCGN3271, respectively. The total protein contents of E. coli containing pCGN3270 (2-1 cDNA) and pCGN3271 (5-2 cDNA) were analyzed by SDS-PAGE (Fig. 4). In each case, a new protein band was observed in the IPTG-induced cultures relative to the pET8c control. Although the computer-predicted molecular masses of the 2-1 and 5-2 encoded proteins are very similar (within 273 D, see above), the mobilities of these proteins as expressed from pCGN3270 and pCGN3271 are significantly different. The protein encoded by pCGN3270 has a molecular mass based on electrophoretic mobility of approximately 40 kD, whereas the molecular mass of the protein encoded by pCGN3271 is approximately 36 kD. The induced proteins indicated in Figure 4 were subjected to N-terminal sequencing to confirm their identities. In each case, the protein sequence matched that predicted by the cDNA (data not shown).

Extracts of cells expressing either pET8c (control), pCGN3270 (2-1), or pCGN3271 (5-2) were assayed for thioesterase activity using 18:1-ACP, the preferred substrate for safflower 18:1-ACP thioesterase (12). As shown in Figure 4, the 18:1-ACP thioesterase activity in cells containing pCGN3270 and pCGN3271 was approximately 100- and 50-fold higher, respectively, than that in control cells.

To characterize further the safflower 18:1-ACP thioesterases, the chain-length specificity of the thioesterase activities expressed from the cDNA clones was compared to that of *E. coli* and a crude safflower seed extract. As shown in Table I, pCGN3270 and pCGN3271 cultures contained thioesterase activity characteristic of safflower seeds, i.e. with a much higher preference for 18:1-ACP over 18:0-ACP. The data also indicate that the two safflower enzymes have slightly different activities on the saturated 18:0-ACP and 16:0-ACP substrates. The enzyme expressed from pCGN3271 shows less discrimination against the saturated acyl-ACP substrates than the enzyme expressed from pCGN3270.

DISCUSSION

This paper presents the first sequencing and molecular characterization of 18:1-ACP thioesterase from a higher plant. Protein purification resulted in the elution of two peaks of thioesterase activity from a chromatofocusing column, each of which contained prominent protein species of 34- and 40-kD. Protein sequence data suggested that all four species were derived from a single protein. Two cDNA clones encoding 18:1-ACP thioesterase have been isolated, one of which corresponds to the purified protein.

There are several published studies of acyl-ACP thioesterases from a variety of plant species (7, 8, 12, 17, 18, 24), and recently the purification and cloning of a 12:0-ACP thioesterase from California bay has been reported (6, 30). We used as a starting point for our purification the protocol described by McKeon and Stumpf (12), but found that in our laboratory it did not yield the degree of purity reported by those authors. This may have been due, among other reasons, to differences in the variety of safflower tissue used for our isolations, or to the variability in the selectivity and elution properties of ACP-Sepharose columns. Therefore, we added a chromato-

5-2							GG	GTAA	CATG	GCAT	AAAC	GTGA	АТАА	CTGC	AACI	CCAC	STGTO	ACTI	TCCC	TTTC	CTTI	CCAC	CACC	ATCT	сст	73
2-1 5-2	ссст	CGG1 GC	CCCF	ACCA	GGAA	алас Алаа	ТССА АААА	TAAA	ACCA	CCAC	CACC	TCTT	TCGO	TCAA CATC	GGAG	TCT	CCG	ACC	CCAC	CACO	ACCA	CCGC	CGCC	GGCA	ACT ATC	172 93
2-1 5-2	MET ATG	Leu CTA	Ser TCA G	Arg CGA 	Pro CCT	Leu CTT AAA	Pro CCG GGT	Thr ACC G.T	Thr ACC C.G	Ala GCC G	Ala GCG A	Ala GCG C	Ala GCG 	Thr ACC GTG	Thr ACG G	Thr ACG G	Thr ACG .T.	Asn AAT T.C	Asn AAT	Cys TGC GC.	Asn AAT TCC	Gly GGC .C.	Val GTC AAA	Asn AAC G	Ser TCC A.T	25 247 162
2-1 5-2	Arg CGC ACT	Gly GGC TTT	Ala GCC	Leu TTA C	Pro CCT A	His CAT	Ser TCC	Arg CGA	Ser TCC	· Val GTT A	Gly GGA	Phe TTC	· Ala GCC .T.	Ser TCG	Ile ATT	Arg CGG	Lys AAA	Arg CGA	Ser AGC TA	Thr ACC	Gly GGT	Ser TCC	Lys Leu TTA	Asp Cys TGC	Thr Asn AAT	23 50 322 237
	Thr Ser	Phe	•	Pro	Thr Pro	Arg	Thr	Val	Ala	Ile Pro	Val	Ser MET	Val <u>SO8</u> 3 Ala	val	Ara	Thr	Arg	Glu	Tyr	Asn	Val Thr	Phe		 Ala	 	48
2-1 5-2	ТСG Т	TCG Ser	TCG Ser	CCG T Ser	CCG T Ser	CGG A.A	ACG .A. Lys	GTG T	GCG T.T Ser	CCG 	GTG T Leu	ATG C.A Leu	GCG	GTG	AGG GC. Ala	ACC	GGT A	GAG 	CAA G	CCG 	ACC .G. Ser	GGC T	GTT	GCC	GTC	391 309 72
2-1	Gly GGA	Leu TTG	Lys AAG	Glu GAG	Ala GCG	Glu GAG	Ala GCG	Glu GAG	Val GTG	Glu GAG	Lys AAG	Ser AGC	Leu CTG	Ala GCG	Asp GAT	Arg CGG	Leu CTT	Arg CGG	MET Atg	<u>SO82</u> Gly GGG	Ser AGC	Leu TTG	Thr ACG	Glu GAA	Asp GAT	98 466
5-1	- Glv	- Leu	Ser	Leu	Arg			Asp	Lys	 Tle		 Cv •		Gly	Asn Val	<u> </u>	· · · ·		Leu	 	· · · ·		···· ·		 The	95
2-1 5-2	GGA	ΤΤG λ	тс G 	TAT 	AAG	GAG	AGG .A. Lys	TTC	ATC G.T Val	ATA 	AGG	TGT	TAT	GAA 	GTC	GGG A	ATT 	AAT C	AAG A	ACT	GCA T	ACT	GTT A Ile	GAA 	ACC G	541 453 120
2-1 5-2	Ile ATT	А1а GCT А	λ sn λλ Τ 	Leu CTA G	Leu TTG	Gln CAG	Glu GAG	Val GTT	Gly GGA	Gly GGT	Asn AAT	His Cat	Ala GCT	Gln CAG	Ser AGT G	Val GTT	Gly GGA	Phe TTT	Ser TCA T	Thr ACA	Азр, GAC Т	Gly GGA G	Phe TTT	Ala GCC	Thr ACC	148 616 528
2-1	Thr ACG	Thr	Thr	MET	Arg	Lys AAA	Leu	His Cat	Leu	Ile	Trp TGG	Val	Thr	Ser	Gly Arg	MET	SO8	34B Ile	Glu	Ile	Tyr	Arg	Tyr	Pro	Ala	145 173
5-2					A.G			· · · ·	· · · · ·	· · · · ·		т	···	G.A Ala	· · ·		T		· · ·	A	T		c	т	•••	603 170
2-1 5-2	Trp TGG 	Ser AGT 	Азр GAT 	Val GTG 	Val GTT A Ile	Glu GAA 	Ile ATC T	Glu GAG 	Thr ACT	Trp TGG 	Cys TGT GT. Val	Gln CAA G	Ser AGT G Glv	Glu GAA G	Gly GGA G	Arg AGG .A. Lvs	Ile ATT G.C Val	Gly GGG	Thr ACT	Arg AGA G	Arg CGT	Asp GAT	Trp TGG	Ile ATT	MET ATG C.C	198 766 678 195
	508	22											-							-	-	-		•		
2-1 5-2	Lys AAA 	Asp GAC	His CAT T TVF	Ala GCG C	Ser AGT .A.	Gly GGT	Glu GAA G	Val GTC T	Ile ATT	G1y GGA	Arg AGG	Ala GCT	Thr ACA	Ser AGC 	Lуз Алл 	Trp TGG	Val GTG	MET Atg	MET Atg	<u>SOB</u> Asn AAC	34 Glu GAG	Asp GAT	Thr ACT	Arg AGA	Arg AGA	223 841 753
2-1 5-2 2-1 5-2	Lys AAA Leu CTC T.G	Asp GAC Gln CAG	His CAT T Tyr Lys AAA	Ala GCG C Val GTC	Ser AGT .A. Asn Asn AAC .GT	Gly GGT Asp GAT	Glu GAA G Asp GAC T	Val GTC T Val GTC	Ile ATT Arg AGA 	Gly GGA Asp GAC G	Arg AGG Glu GAA G	Ala GCT C Tyr TAT	Thr ACA Leu CTC T.A	Ser AGC Val GTT G	Lys AAA Phe TTT	Trp TGG Cys TGT C	Val GTG Pro CCC	MET ATG Lys AAG .G.	MET ATG Thr ACA	SOB Asn AAC Pro CCA TTG	Glu GAG Arg AGA	Asp GAT Leu TTA	Thr ACT Ala GCA	Arg AGA Phe TTT	Arg AGA Pro CCT	223 841 753 220 248 916 828
2-1 5-2 2-1 5-2 2-1	Lys AAA Leu CTC T.G Glu GAA	Asp GAC Gln CAG Lys AAG	His CAT T Tyr Lys AAA Asn AAC	Ala GCG C Val GTC Thr ACT	Ser AGT .A. Asn Asn AAC .GT Ser Ser AGC	Gly GGT Asp GAT Ser AGC	Glu GAA G GAC T Leu CTG	Val GTC T Val GTC Lys AAG	Ile ATT Arg AGA Lys AAA	Gly GGA Asp GAC G Glu Ile ATA	Arg AGG Glu GAA G Ala GCA	Ala GCT C Tyr TAT Lys AAA	Thr ACA Leu CTC T.A Leu CTA	Ser AGC Val GTT G Glu GAA	Lys AAA Phe TTT Asp GAC	Trp TGG Cys TGT C Pro CCC	Val GTG Pro CCC Ala GCC	MET ATG Lys AAG .G. Arg Glu GAA	MET ATG Thr ACA Tyr TAT	SOB Asn AAC Pro CCA TTG Leu Ser TCG	Glu GAG Arg AGA Thr ACG	Asp GAT Leu TTA Leu CTA	Thr ACT Ala GCA Gly GGG	Arg AGA Phe TTT Leu CTT	Arg AGA Pro CCT Val GTG	223 841 753 220 248 916 828 245 273 991
2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Leu CTC T.G Glu GAA 	Asp GAC Gln CAG Lys AAG G Glu	His CAT T Tyr Lys AAA Asn AAC	Ala GCG C Val GTC Thr ACT .AC Asn	Ser AGT .A. Asn Asn ASn AC .GT Ser AGC .AT Asn	Gly GGT Asp GAT Ser AGC 	Glu GAA G GAC T Leu CTG A MET	Val GTC T Val GTC Lys AAG 	Ile ATT Arg AGA Lys AAA Sog	Gly GGA Asp GAC G Glu Ile ATA 	Arg AGG Glu GAA G Ala GCA C Pro	Ala GCT C Tyr TAT Lys AAA 	Thr ACA Leu CTC T.A Leu CTA G	Ser AGC Val GTT G Glu GAA 	Lys AAA Phe TTT Asp GAC T	Trp TGG Cys TGT C Pro CCC A	Val GTG Pro CCC Ala GCC T	MET ATG Lys AAG .G. Arg Glu GAA 	MET ATG Thr ACA Tyr TAT 	SOB Asn AAC Pro CCA TTG Leu Ser TCG C	Glu GAG Arg AGA Thr ACG .G. Arg	Азр GAT Leu TTA Leu CTA T	Thr ACT Ala GCA Gly GGG A	Arg AGA Phe TTT Leu CTT 	Arg AGA Pro CCT Val GTG 	223 841 753 220 248 916 828 245 273 991 903 270
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Leu CTC T.G Glu GAA Pro CCA	Asp GAC Gln CAG Lys AAG G Glu Arg AGA G	His CAT T Tyr Lys AAA Asn AAC Asn AAC 	Ala GCG C Val GTC Thr ACT .ACT Asn Ala GCC T Ser	Ser AGT . A. Asn Asn Asn Asc . GT Ser AGC . AT Asn GAT 	Gly GGT Asp GAT Ser AGC Leu CTC T.G	Glu GAA G GAC T Leu CTG A MET Asp GAT 	Val GTC T Val GTC AAG AAG 	Ile ATT AGA Lys AAA SOB AAA 	Gly GGA GAC G Glu Ile ATA Lys AAG A	Arg AGG Glu GAA G Ala GCA C Pro His CAT C	Ala GCT C Tyr TAT Lys AAA Val GTT	Thr ACA Leu CTC T.A Leu CTA G Asn AAC	Ser AGC GTT G Glu GAA Asn AAT 	Lys AAA Phe TTT GAC T Val GTT 	Trp TGG TGT C Pro CCC A Thr ACC	Val GTG Pro CCC Ala GCC T Tyr TAC	MET ATG AAG .G. Arg Glu GAA Ile ATT C	MET ATG Thr ACA Tyr TAT Gly GGA G	SOB Asn AAC Pro CCA TTG Leu Ser TCG C Trp TGG 	Glu GAG Arg AGA Thr ACG .G. Arg Val GTT .C. Ala	Asp GAT Leu TTA Leu CTA T A	Thr ACT Gly GGG A Glu GAG 	Arg AGA Phe TTT CTT Ser AGC 	Arg AGA Pro CCT Val GTG Ile ATC 	223 841 753 220 248 916 828 245 273 991 903 270 298 1066 978 295
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Leu CTC T.G Glu GAA Pro CCA 	Asp GAC Gln CAG Lys AAG G Glu AGA Gln CAA CA 	His CAT T Tyr Lys AAA Asn AAC Glu GAA 	Ala GCG C Val GTC ACT .ACT Asn Ala GCC T Ser Val	Ser AGT Asn Asn Asn Asc Ser Ser AGC Asn Asp GAT 	Gly GGT Asp GAT Ser AGC Leu CTC T.G GAC 	Glu GAA G Asp GAC T Leu CTG A MET Asp GAT Thr AcT C	Val GTC T Val GTC Lys AAG His CAT	Ile ATT Arg AGA Lys AAA SOB Asn AAC Glu GAA 	Gly GGA Asp GAC G Glu Ile ATA 29 Lys AAG A Leu CTA G	Arg AGG Glu GAA G Ala GCA C Pro His CAT C Gln CAA 	Ala GCT C Tyr TAT Lys AAA Val GTT Thr ACG G.T Ala	Thr ACA Leu CTC T.A Leu CTA G Asn AAC Ile ATT	Ser AGC GTT G Glu GAA Asn AAT Thr ACC	Lys AAA Phe TTT Asp GAC T Val GTT Leu CTA	Trp TGG Cys TGT C Pro CCCC A Thr Acc Asp GAC	Val GTG Pro CCC Ala GCC T Tyr TAC 	MET ATG Lys AAG .G. Arg Glu GAA Arg CGG CGG A.A	MET ATG Thr ACA Tyr TAT Gly GGA G Arg CGG	SOB Asn AAC Pro CCA TTG Leu Ser TCG C TTp TGG C Glu GAA	34 Glu GAG Arg AGA Thr ACG .G. Arg Val GTT .C. Ala Cys TGC	Asp GAT Leu TTA Leu CTA T .A Gln CAG	Thr ACT Ala GCA Gly GGG AC GAG His CAT GG	Arg AGA Phe TTT Leu CTT Ser AGC Asp GAC 	Arg AGA Val GTG Asp GAC	223 841 753 220 248 916 828 245 273 991 903 270 298 1066 978 295 323 1141 1053 320
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Leu CTC T.G Glu GAA Pro CCA Ile ATA 	Asp GAC GAC CAG CAG CAG CAG CAG CAG CGLU Arg AGA CGLU CAA .C. Pro Val GTC T	His CAT T Tyr Lys AAA AAC Aan AAC Glu GAA Asp GAT	Ala GCG C Val GTC ACT .ACT ACT .ACT ACT .ACT Ser Val GTC A Ile Ser TCC A	Ser AGT ASI ASI ASI ASI ASI ASI ASI ASI ASI ASI	Gly GGT Asp GAT Ser AGC Leu CTC T.G GAC Thr ACG 	Glu GAA G GAC T Leu CTG A MET Asp GAT C Ser AGT C	Val GTC T Val GTC Lys AAG His CAT 	Ile ATT AGA AGA SOB AGA AC Glu GAG GAG CGT	Gly GGA GGA GAC GAC GAC GAC Lys AAG CTA CTA Ser TCA GA.	Arg AGG Glu GAA G Ala GCA C Pro His CAT C Gln CAA Leu CTA .C.	Ala GCT GCT Tyr TAT Lys AAA Val GTT Thr ACG G.T Ala Leu CTC 	Thr ACA Leu CTC T. A Leu CTA Asn AAC Asn AAC Asp GAC G	Ser AGC · Val GTT G Glu GAA · · · · · · · · · · · · · · · · · ·	Lys AAA Phe TTT Asp GAC T Val GTT Leu CTA T Ala GCC T	Trpg TGG Cys TGT Pro CCC A Thr ACC Asp GAC Ala GCC A	Val GTG Pro CCC Ala GCC T Tyr TAC S Tyr TAC Ile ATC GGT	MET ATG ATG .G. Arg Glu GAA Ile ATT C Arg CGG A.A Ser TCG GTC	MET ATG Thr ACA Tyr TAT Gly GGA G Arg CGG T Lys AAA	SOB Asn AAC ···· Pro CCA TTG Leu Ser TCG · TTG Glu GAA ···· · ·	34 Glu GAG A Arg AGA Thr AGA .G. Arg Val GTT .C. Ala Cys TGC Glu GAA A	Asp GAT Leu TTA Leu CTA GIN CAG Gly GGA .A.	Thr ACT Ala GCA GGG CA CGG CAT CAT CAT CAT CAT CAT CAT CAT CAT CAT	Arg AGA Phe TTT Leu CTT Ser ASC Asp GAC Asn AAC	Arg AGA	223 841 753 220 248 828 245 273 991 293 270 298 295 323 1141 1053 320 348 2125
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Leu CTC T. G Glu GAA Pro CCA Ile ATA Ser TCT	Asp GAC Gln CAG Lys AAG G Glu Arg AGA Pro Val GTC GT CT	His CAT T Tyr Lys AAA Asn AAC Asn AGA Yal GAT Yal	Ala GCG GCG C Val GTC Asn Ala GCC T Ser Val GTC A Ile Ser TCC A	Ser AGT Asn Asn Asn Ser Ser AGC Asn Asp GAT Ile ATC Leu CTC 	Gly GGT GGT GGT Asp GAT Ser AGC CTC GAC Thr AASP GAC Leu CTC T.G GAC Lu Sar Asp CTC T.G ASP 	Glu GAA G Asp GAC T Leu CTG A MET MET MET MET C Ser ACT C Asp GAC	Val GTC T Val GTC Lys AAG His CAT CAT CAT CAT Glu GAA	Ile ATT Arg AGA Lys AAA A SOB AAA A Glu GAA CGT Arg Thr ACG	Gly GGA GGA GGL Ile ATA Lys AAG CTA GLU CTA GLU CTA GLU CTA GLU ASP GAT	Arg AGG Glu GAA G Ala GCA C Pro His CAT C Gln CAA C Pro Leu TTG	Ala GCT C Tyr TAT Lys AAA Val GTT Thr ACG G.T Ala Leu CTC Ser AGC	Thr ACA CTC T.A Leu CTC T.A Leu CTA A Sn AASN AASN AASN AASN GAC GAC GAC GAC	Ser AGC GTT G Glu GAA Asn AAT Asn AAT Asp GAT A Asn Phe TTT	Lys AAA Phe TTT Asp GAC Val GTT Ala GCC T Leu TTG	Trp TGG TGT C Pro CCC A Thr ACC Asp GAC Ala SCC A His CAT	Val GTG Pro CCC Ala GCC Tyr TAC Tyr TAC GGT Gly Leu TTA	MET ATG ATG .G. ATG Glu GAA Ile ATT CGG A.A Ser TCG GTC Val Leu CTA	MET ATG Thr ACA Tyr TAT Gly GGA Gly Gly Gly Gly Arg CGG AAAA AAA AAA AAA AAAAA AAAAA AAAAA AAAAA AAAAAA	SOB Asn AAC Pro CCA TTG Leu Ser TCG Trp TCG 	34 Glu GAG AGA Arg AGA Thr ACG .G. Arg Val GTT .C. Ala Cys TGC Glu GAA A Lys Ser TCC	Asp GAT Leu TTA T Leu CTA T Gln Gln GGA A. Glu Gly GGC	Thr ACT ALT GCA GCA GCA GCA GCA GCA GCA CA CA CA CA CA CA CA CA CA CA CA CA C	Arg AGA AGA Phe TTT Leu CTT Ser AGC Asp GAC Asn AAC CT Asp GAC CT 	Arg AGA Pro CCT Val GTG ATC Gly GGA Leu CTC	223 841 753 220 228 916 828 245 273 991 903 270 298 1066 978 295 323 1141 1053 320 348 1216 1125 344
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAAA Leu GTC T.G Glu GAA Pro CCA Pro CCA Ile ATA Ser TCT Glu	Asp GAC GING CAG CAC GING AAG C GIU AAG C GIU AGA C. Pro Val GTC C Yal Leus Ser TCT GT. 	His CAT T Tyr Lys AAA Asn Asn Glu GAA Val GTT TCC Ser Asn	Ala GCG C Val GTC AcT .AC Asn Ala GCC T Ser Val GTC A Ile Ser TCC A	Ser AGT AAT AAT AAT AAT AAT AAT AAT AAT AAT	Gly GGT GGT ··· Ser Asp GAT ··· Ser AGC ··· ASP GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· Asp GAT ··· AGC ··· Asp CAT ··· Asp CAT ··· AGC ··· Asp CAT ··· AGC ··· Asp CAT ··· AGC ··· Asp CAT ··· AGC ··· Asp CAT ··· ASP ··· ASP ··· Asp ···· Asp ···· Asp ···· Asp ···· Asp ···· Asp ···· ··· Asp ···· ··· ··· Asp ···· ··· ··· ··· ··· ··· ··· ··· ···	Glu GAA G Asp GAC T Leu CTG A MET ASP GAT C Ser ACT C Sag C Ser ASP C C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C C G C C G C C G C C G C C G C C G C C G C C G C C C G C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C G C C C C G C	Val GTC T Val GTC Lys AAG His CAT Glu GAA Glu	Ile ATT Arg AGA SOB: AAA SOB: AAA AAA Glu GAA CGT Arg Thr ACG CAA Gln Trp	Gly GGA GGA GGA GGA GGA C GGA C GGA GGA C Asp GAT C Asp GAT	Arg AGG Glu GAA G C Pro His CAT C Gln CAA Pro Leu TTG C.A Lys	Ala GCT C Tyr TAT Lys AAA Val GTT	Thr ACA Leu CTC T.A GTA Asn AAC Asn AAC GAC .GA Gly Arg CGG C.A Pro	Ser AGC AGC Val GTT G Glu GAA Asn AAT Asp GAT Asp GAT Asp GAT Asp GAT Asp AGC 	Lys AAA Phe TTT Asp GAC T Val GTT Leu TTG A MET Lys	Trp TGG Proc CCC A Thr ACC Asp GAC His CCT Lys	Val GTG Pro CCC Ala GCC Tyr TAC Ile GGT Gly Leu TTA C	MET ATG Lys AAG .G. Arg Glu GAA Arg CGA A.T C	MET ATG Thr ACA Gly GGA CGG CGG Lys AAA G Arg CCGG 	SOBA Asn AAC Proc CCA TTG Leu Ser TCG C TTC TCG C.C	34 Glu GAG GAG AGA AGA AG CA GL GL GL GL GL GL GL GL GL GL GL GL GL	Asp GAT Leu TTA T CTA T Leu CTT A Gln Gly GGA Gly GGC 	Thr AACT AACT ACT ACT ACT ACT ACT ACT ACT A	Arg AGA AGA Phe TTT CTT Ser AGC Asp GAC Asp GAC CTT 	Arg AGA	223 841 753 220 248 916 828 245 273 991 903 270 270 298 1066 978 295 323 323 323 344 373 1290 369 389
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lya AAA Leu CTC T.G Glu GAA Pro CCA Pro CCA Glu GAA Glu GAA Glu GAA 	Asp GAC Gln CAG C.G Glu AAGA Glu AAGA Glu CAA Glu CAA Glu CAA C. C. Yal CTC TCT GT. Val Leu CTA A.C GTC	His CAT T Tyr Lys AAA Asn Asn Asn Asn Glu GAT Val GTT TCC Ser Asn AAT C	Ala GCG C Val GTC Ac Asn Ala GCC T Ser TCC Arg AGG 	Ser AGT Asn Asn Asn Asn Ser AGC AT Asn GAT Ile ATC GI y GGT T Cys	Gly GGT Asp GAT Ser AGC Leu CTC T.G GAC Thr AASP GAC Asp GAC Asp CCC Asp CCC 	Glu GAA G GAC T Leu CTG GAC T MET Asp GAT C Ser ACT C	Val GTC T Val GTC ATG His CAT Glu GAG CAC Glu GAG CAC 	Ile ATT ATT AGA AGA SOB AGA AGA AGA AGA AGA AGA AGA AGA AGA AG	Gly GGA Asp GAC G Glu Ile ATA C Glu Lys AAG A Ser TCA GA. Glu Asp GAT AGA Glu Asp GAC G CACAC G CAC C CAC G CAC G CAC CAC 	Arg AGG Glu GAA G Ala GCA C Pro His CAT C Gln CAA CTA .C Pro Leu TTG C.A	Ala GCT C Tyr TAT Lys AAA Yal GTT Thr ACG G.T Ser AGC Leu Ser AAA 	Thr ACA ACA CTCC T.A Leu CTCC T.A	Ser AGC Val GTT G GIU GAA Asn AAT A Asn Phe TTT Ala GCG A	Lys AAA Phe TTT Asp GAC T Ual GTT Ala GCC T Leu TTG AAA 	Trp TGG Cys TGT C Pro CCC A Thr ACC Asp GAC Ala GCC Ala SCC Ala SCC Ala AA AAA AAA AA AAA AAA AA	Val GTG Proc Ccc Ala GCC Tyr TAC Ile GGT Gly Leu TTA C STGA .A.	MET ATG Lys AAG .G. Arg Glu GAA Arg CGG A.A Ser TCG GTC Val Leu CTA GCA GCA	MET ATG Thr ACA Glyy GGA G GGG Lys AAA A Arg CGA A Arg CGA 	SOBL Asn Asn Acc CCA TTG Leu Ser TCG CCA TTG Glu Glu GAA CTC TTT Phe Ser TCA CCTT TCA CCTT TCA	34 Glu GAG GAG AGA AGA AGA Thr ACG .G. Arg GTT .C. Ala GTT .C. Ala GLU GAA A Cys TGC Glu GAA GTT .C. ACG Cys TGC C. C. A CGC GCC GCC GCC GCC GCC GCC GCC GCC G	Asp GAT Leu TTA Leu CTA GIT CAG CAG Gly GGA Gly GGC TTGT	Thr ACT ACT ACT ACT ACT ACT ACT ACT ACT ACT	Arg AGA AGA Phe TTT CTT Asp GAC Asp GAC Gly GGT CCTT 	Arg Arg AGA Val GGG	223 841 753 220 248 916 828 245 273 991 903 270 298 295 323 1141 1053 320 348 1216 1053 320 348 1226 344 373 1291 1200 369
2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2 2-1 5-2	Lys AAA Gluu CTC T.G Glu GAA Pro CCA Ile CCA Ser TCT Glu GAA 	Asp GAC Gln CAG G Glu AGA Glu AGA Glu AGA C C Val CTA CTA CTA CTA CTA CTA CTA CTA CTA CTA	His CAT T Tyr Lys AAA Asn Asn Acc Asn Asn Cat Cat Cat Cat Cat Cat Cat Cat Cat Cat	Ala GCG Val GTC ACT .ACT ACT .ACT ACT .ACT ACT ACT .CC ACT Ser Val GTC CAAA CC ACT CCC ACT ACT CCC CCC ACT CCC CCC ACT CCC CCC CCC	Ser AgT Asn Asn Asn Ser Ser AgC C.AT Asn Asp GAT Ile CTC C Gly GGT T Cys AAAC CTC	Gly GGT Ser ASC GAT Ser ASC CTC T.G CTC T.G CTC T.G ASP GAC Asp GAC Asp CTC T.G CTC T.G CTC CTC T.G CTC CTC T.G CTC 	Glu GAA G GAC T GAC CTG A MET Asp GAC Ser AGT C Thr ACT ACC Thr ACC CTG A Thr ACC CTG A Thr ACC CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG A Thr CTG C Thr CCG C Thr CCG C Thr CCG C Thr CCG C Thr CCG C Thr CCG C Thr CCG C Thr CCC C Thr CCC C Thr CCC C Thr CCC C Thr CCC C Thr CCC C Thr CCC C CCC C CCC C CCC C CCC C CCCC C CCCCCC	Val GTC Val GTC Lys AAG C His CAT CCT Glu GAA Glu GAG GAA CAT TTTTC	Ile ATT Arg AGA AGA SOBJ AGA C Glu GGA CGT Arg Thr ACGG CGT Thr ACGG CGT Thr CGGC CGT TTCGG	Gly GGA Asp GGC GGL Ile ATA ASP CTA CTA GGL ASP GGA CTA CTA GGA CTA CTA CTA CTA CTA CTA CTA CTA CTA CT	Arg AGG AGG GAA G GAA G Pro His CAT C Pro Leu TTAG C.A Lys AAG TTAG CCTA	Ala GCT C Tyr TAT TAT Val GTT Val GTT AACA G.T Ala Leu CTC Ser AACA G GTTT TTTG	Thr ACA Leu CTC T.A Leu CTA Gac Asn AAC	Ser AGC Val GTT GTU GAT ASN ASN ASN ASN ASN ASN ASN ASN ASN ASN	Lys AAA TTT Phe TTT Asp GAC T Val GTT Ala GCC T Leu TTG A.AA AC A.AA T TTTA	Trp TGG Cys TGT Tor ACC Asp GAC Asp GAC Asp GAC Asp GAC Asp GAC Asp GAC Asp GAC Asp GCC Asp GCC Asp GCC Asp GCC Asp GCC Asp GAC Asp GAC Asp GAC Asp GAC Asp GCC Asp GAC Asp GAC Asp GAC Asp GAC Asp GAC Asp GCC Asp GAC Asp GCC A ACD A A A A A A A A A A A A A A A A A	Val GTG CCC CCC Ala GCC Tyr TAC Ile GGT Gly Leu TTA C SGGT	MET ATG ATG Jus Ang Glu GAA Ile ATT T.CCGG A.A Ser TCCG GTC Val Leu CTA GCA GCA GCA	MET ATG Thr ACA Tyr TAT TAT TAT CGG GGA T CGG CGA AArg CCGA A Arg CCGA A Arg CCGA A	SOBE Asn Asn Arc CCA TTG CCA TTG CCA TTG CCA TTG Glu CTC TTT GA CTC TTT Phe Ser TCA CCT TTAGAT	Arg Glu GAG AGA AGA AGA AGA AGA AGA AGA AGA AG	Asp GAT TTA Leu CTA T Glu GCTA Glu GGC CTTGT CCTAG	Thr ACT Ala GCA GCA GCA GCA GCA GAG Arg CAT CAT CAT ACC TAC Asp GAG AS F TAG GAG CAT Thr TACG GAG ACT CAT CAT CAT CAT CAT CAT CAT CAT CAT	Arg Arg AGA Phe TTT TTT Ser AGC Asp GAC Gly GGT CGTAC CCGT	Arg Arg AGA . Pro CCT . Yal GGTG .	223 841 753 220 248 916 828 245 273 991 903 270 298 295 323 1141 1053 320 348 1216 1125 334 348 1216 1125 344 373 1291 1200 369 369 1373 1282 385

Figure 3. Nucleotide and deduced amino acid sequences of safflower 18:1-ACP thioesterase cDNA clones 2-1 and 5-2. DNA sequences of both clones are shown in the 5'- and 3'-untranslated regions. Throughout the coding region, only the complete sequence of 2-1 is displayed. Sequence of 5-2 is shown where it differs from 2-1. Dots indicate identity to 2-1; dashes indicate gaps introduced to preserve the alignment. The location of the sequenced CNBr peptides are indicated by overlining of the 2-1 amino acid sequence. Arrowheads indicate the transit peptide cleavage sites. A consensus polyadenylation signal (19) is underlined in both the 2-1 and 5-2 sequences.

KNUTZON ET AL.



Figure 4. Expression of safflower 18:1-ACP thioesterases in *E. coli.* IPTG-induced *E. coli* cells expressing a control plasmid (pET8c), safflower 2-1 cDNA sequences (pCGN3270), or safflower 5-2 cDNA sequences (pCGN3271) were analyzed by SDS-PAGE. Total protein (10 μ g) from cell lysates was electrophoresed through a 10% Trisglycine gel and stained with Coomassie brilliant blue. Positions of molecular mass standards (in kD) are indicated on the right. Extracts of *E. coli* containing either pET8c, pCGN3270, or pCGN3271 were assayed for thioesterase activity using a [1-¹⁴C]18:1-ACP substrate. Activity is expressed as microunits (μ U) per μ g of protein. One microunit of enzyme catalyzes the hydrolysis of one pmol of substrate per min at 30°C.

focusing step to obtain material suitable for protein sequencing. However, even after this additional step there were still several proteins present in the two thioesterase activity pools, the predominant ones having molecular masses of 40 and 34 kD. Safflower 18:1-ACP thioesterase as purified by McKeon and Stumpf (12) had a reported subunit mol wt of 41,000. However, the purified California bay 12:0-ACP thioesterase has a mol wt of 34,000 (6). Due to the prominence of both the 34- and 40-kD proteins in our preparation, we chose both for sequence analysis.

Protein sequence analysis of the 34- and 40-kD proteins found in each activity pool indicates that these isoforms are likely derived from a single protein. No differences in protein sequence of internal CNBr fragments of any of the four species were observed and the amino-terminal sequences of the 34- and 40-kD proteins in the second activity pool were identical. We suggest that the different mol wt species of 18:1-ACP thioesterase result from the removal of a C-terminal peptide either by processing in vivo or by degradation during the extraction and purification. It is not likely that the during the acid precipitation step resulting in a change in pI, or there may be other modifications to the protein as yet

undiscovered. Although the protein sequence evidence suggests that all of the thioesterase protein species observed in our partially purified fractions may be isoforms of a single protein, two highly homologous but distinct classes of cDNAs were isolated from our safflower embryo cDNA library. All of the protein sequences that were determined may be deduced from the translated 2-1 cDNA sequence, suggesting that it encodes the proteins that we purified. No peptides were found that matched the 5-2 cDNA clone with 100% identity. Not only is the N-terminal sequence encoded by the 5-2 clone different than that from the 2-1 clone, but the size of the CNBr fragments obtained from this protein would also be different. It is possible that the protein encoded by clone 5-2 is much lower in abundance and that it was not a sufficiently prominent band to be considered for sequencing. As a corollary to this argument, it is possible that the protein was only a minor component of the digested sample, and that the CNBr fragments were not abundant enough to be detected after SDS-PAGE and electroblotting. Alternatively, examination of the predicted pIs of the two protein products indicates that 5-2 encodes a much more basic protein than does 2-1. Thus, the protein corresponding to 5-2 may have been eliminated during the acid precipitation step of protein purification.

Both classes of cDNA clones encode 18:1-ACP thioesterase activity when expressed in *E. coli*, but the activities have slightly different substrate specificities. The protein encoded by 5-2 has an apparently lower molecular mass than that encoded by 2-1, as determined by SDS-PAGE, although the computer-predicted mol wts of the encoded proteins are similar. (Note that this is not the same mol wt difference as that observed between the two protein isoforms during protein isolation.) The apparent mass difference between the

Table I. Acyl Chain Length Specificity of Safflower Thioesterase

 Produced in E. coli

Crude extracts of *E. coli* cells containing pET8c, pCGN3270 (2-1 cDNA), or pCGN3271 (5-2 cDNA) and a crude safflower seed extract were assayed for thioesterase activity on a series of [¹⁴C]acyl-ACP substrates. The activities on each substrate are normalized to the activity on 18:1-ACP (=100). Values represent the mean \pm range from two separate induced cultures or two separate safflower seed extracts.

Extract	Acyl-ACP Substrate											
LAUACI	12:0	14:0	16:0	18:0	18:1							
	% of activity on 18:1 ACP											
pET8c	28.1 ± 2.3	22.1 ± 1.7	53.5 ± 0.6	104.9 ± 0.1	100							
pCGN3270	0.5 ± 0.3	0.6 ± 0.1	9.3 ± 0.8	9.7 ± 1.2	100							
pCGN3271	1.2 ± 0.3	1.3 ± 0.3	23.5 ± 0.2	29.9 ± 0.3	100							
Safflower	5.0 ± 0.3	2.2 ± 0.2	14.7 ± 0.1	15.3 ± 1.8	100							



encoded proteins expressed in *E. coli* raises the possibility that the two thioesterases undergo different posttranslational processing and that this may contribute to the slight difference in substrate specificity observed between the two proteins. The role that the two thioesterases play in plants is unknown, but multiple isoforms of other fatty acid biosynthetic enzymes have been characterized in a number of plant species (4). Perhaps one thioesterase is specific to embryo tissue and is involved in storage oil synthesis, whereas the other serves a more general function in structural lipid formation. Further characterization of these functions will be the focus of future investigations.

The amino acid sequence of the mature safflower 18:1-ACP thioesterase encoded by clone 2-1 was compared to the deduced amino acid sequences of an acyl-ACP thioesterase from *Brassica rapa* (J.C. Kridl, unpublished data) and a 12:0-ACP thioesterase from California bay (30). As shown in Figure 5, there is a high degree of homology (74% identity) between the safflower and *Brassica rapa* clones, and a lesser, but significant, homology (31% identity) between the safflower 18:1-ACP thioesterase and the 12:0-ACP thioesterase from California bay. As we identify other thioesterases with different chain-length specificities, the regions of sequence involved in catalytic activity and substrate recognition may become apparent.

Neither of the two deduced safflower thioesterase protein sequences shows significant homology to the animal or yeast thioesterases (2, 5, 9), nor do they contain the serine active site consensus sequence (GXSXG) typical of serine proteases, other thioesterases, and acyl transferases (1). The safflower thioesterase activities expressed in E. coli are inhibited by the sulfhydryl-specific reagent N-ethylmaleimide (D.S. Knutzon, unpublished data) and studies indicate that other higher plant acyl-ACP thioesterases are sensitive to N-ethylmaleimide, iodoacetamide, and 5,5'-dithiobis-(2-nitrobenzoic acid) (17, 18), suggesting that a cysteine residue is involved at the active site. There are three cysteine residues conserved between the 2-1 and 5-2 encoded proteins, although none has the conserved active site motif (GXC followed by a hydrophobic residue) of the cysteine-based hydrolases (1). One of the three conserved cysteine residues, however, is also conserved in the California bay 12:0-ACP thioesterase. Labeling studies could verify which of the cysteine residues is the active site of the safflower thioesterases.

The chemical and sequence data indicate that plants have a class of thioesterase distinct from that found in animal **Figure 5.** Dot matrix comparison of deduced amino acid sequences of acyl-ACP thioesterases. The complete deduced amino acid sequences (including transit peptides) were compared using the Pustell Protein Matrix program of MacVector, with a window size of 8 and a minimum percent score of 60. A, Comparison of safflower 2-1 18:1-ACP thioesterase with an acyl-ACP thioesterase cDNA isolated from *Brassica rapa* (J.C. Kridl, unpublished data). B, Comparison of safflower 2-1 18:1-ACP thioesterase with a 12:0-ACP thioesterase cDNA isolated from California bay (30).

systems. The structural differences between plant and animal thioesterases may be related to the requirements for their interactions with their respective fatty acid synthetase complexes. In animals, the thioesterase is a domain of the multifunctional fatty acid synthetase polypeptide (13, 26), whereas in plants the fatty acid biosynthetic enzymes are discrete proteins and it is unknown whether they are associated with one another. The availability of cloned 18:1-ACP thioesterase will allow the study of these interactions as well as further study of the role 18:1-ACP thioesterases play in the determination of seed oil composition.

ACKNOWLEDGMENTS

We gratefully acknowledge William E. Schreckengost for DNA sequencing, Eric R. Svee for substrate preparation, and H. Maelor Davies, C.K. Shewmaker, and Vic C. Knauf for critical reading of the manuscript.

LITERATURE CITED

- Aitken A (1990) Identification of active site esters. In A Aitken, ed, Identification of Protein Consensus Sequences. Ellis Horwood, London, pp 81–91
- Amy CM, Witkowski A, Naggert J, Williams B, Randhawa Z, Smith S (1989) Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. Proc Natl Acad Sci USA 86: 3114–3118
- Blum H, Beier H, Gross HJ (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis 8: 93–99
- Browse J, Somerville C (1991) Glycerolipid synthesis: biochemistry and regulation. Annu Rev Plant Physiol Plant Mol Biol 42: 467–506
- 5. Chirala SS, Kuziora MA, Spector DM, Wakil SJ (1987) Complementation of mutations and nucleotide sequence of *FAS1* gene encoding β subunit of yeast fatty acid synthase. J Biol Chem 262: 4231-4240
- Davies HM, Anderson L, Fan C, Hawkins DJ (1991) Developmental induction, purification, and further characterization of 12:0-ACP thioesterase from immature cotyledons of Umbellularia californica. Arch Biochem Biophys 290: 37-45
- Hellyer A, Slabas A (1990) Acyl-[acyl-carrier protein] thioesterase from oil seed rape: purification and characterization. *In* PJ Quinn, JL Harwood, eds, Plant Lipid Biochemistry, Structure and Utilization. Portland Press Limited, London, pp 157-159
- Imai H, Nishida I, Murata N (1990) Acyl-[acyl-carrier protein] hydrolase of squash cotyledons: purification and characterization. In PJ Quinn, JL Harwood, eds, Plant Lipid Biochemistry, Structure and Utilization. Portland Press Limited, London, pp 160–162
- Kasturi R, Chirala S, Pazirandeh M, Wakil SJ (1988) Characterization of a genomic cDNA clone coding for the thioesterase

domain and 3' noncoding region of the chicken liver fatty acid synthase gene. Biochemistry 27: 7778–7785

- Kunkel TA, Roberts JD, Zakour ŘA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154: 367–383
- Lutcke HÅ, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. EMBO J 6: 43–48
- McKeon TA, Stumpf PK (1982) Purification and characterization of the stearoyl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflower. J Biol Chem 257: 12141–12147
- Naggert J, Witkowski A, Mikkelsen J, Smith S (1988) Molecular cloning and sequencing of a cDNA encoding the thioesterase domain of the rat fatty acid synthetase. J Biol Chem 263: 1146-1150
- 14. Nakamura Y, Yamada M (1974) Long chain fatty acid synthesis in developing castor bean seeds I. The operation of the path from acetate to long chain fatty acids in a subcellular particulate system. Plant Cell Physiol 15: 37-48
- Ohlrogge JB (1987) Biochemistry of plant acyl carrier proteins. In PK Stumpf, ed, The Biochemistry of Plants, Vol 9. Academic Press, Orlando, FL, pp 137–157
 Ohlrogge JB, Kuhn DN, Stumpf PK (1979) Subcellular localiza-
- Ohlrogge JB, Kuhn DN, Stumpf PK (1979) Subcellular localization of acyl carrier protein in leaf protoplasts of Spinacia oleracea. Proc Natl Acad Sci USA 76: 1194–1198
- Ohlrogge JB, Shine WE, Stumpf PK (1978) Fat metabolism in higher plants; characterization of plant acyl-ACP and acyl-CoA hydrolases. Arch Biochem Biophys 189: 382-391
- Pollard MR, Anderson L, Fan C, Hawkins DJ, Davies HM (1991) A specific acyl-ACP thioesterase implicated in mediumchain fatty acid production in immature cotyledons of *Umbellularia californica*. Arch Biochem Biophys 284: 306–312
- Proudfoot NJ, Brownlee GG (1976) 3' noncoding region sequences in eukaryotic messenger RNA. Nature 263: 211–214
- Rock CO, Cronan JE (1981) Acyl carrier protein from Escherichia coli. Methods Enzymol 72: 341–351
- 21. Rock CO, Garwin JL, Cronan JE (1981) Preparative enzymatic

synthesis of acyl-acyl carrier protein. Methods Enzymol 72: 397-403

- Sanger F, Nicklen S, Coulsen AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
- 23. Shimakata T, Stumpf PK (1982) The procaryotic nature of the fatty acid synthetase of developing *Carthamus tinctorius* L. (safflower) seeds. Arch Biochem Biophys 217: 144–154
- Shine WE, Mancha M, Stumpf PK (1976) The function of acyl thioesterases in the metabolism of acyl-CoA's and acyl-acyl carrier proteins. Arch Biochem Biophys 172: 110-116
- 25. Simcox PD, Reid EE, Canvin DT, Dennis DT (1977) Enzymes of the glycolytic and pentose phosphate pathways in proplastids from the developing endosperm of *Ricinus communis*. Plant Physiol 59: 1128-1132
- Smith S, Ágradi E, Libertini L, Dileepan KN (1976) Specific release of the thioesterase component of the fatty acid synthetase multienzyme complex by limited trypsinization. Proc Natl Acad Sci USA 73: 1184–1188
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorft JW (1990) Use of T7 RNA polymerase to direct the expression of cloned genes. Methods Enzymol 185: 60–89
- Stymne S, Stobart AK (1987) Triacylglycerol biosynthesis. In PK Stumpf, ed, The Biochemistry of Plants, Academic Press, Vol 9. Orlando, FL, pp 175-214
- 29. Thompson GA, Scherer DE, Foxall-Van Aken S, Kenny JW, Young HL, Shintani DK, Kridl JC, Knauf VC (1991) Primary structures of the precursor and mature forms of stearoyl acyl carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. Proc Natl Acad Sci USA 88: 2578-2582
- Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Hawkins DJ, Radke SE, Davies HM (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257: 72–74
- Weaire PJ, Kekwick RGO (1975) The fractionation of the fatty acid synthetase activities of avocado mesocarp plastids. Biochem J 146: 425-437