The complete amino acid sequence of a pea (*Pisum sativum*) seed lipoxygenase predicted from a near full-length cDNA

Paul M. EALING* and Rod CASEY

The John Innes Institute and A.F.R.C. Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH, U.K.

A near full-length cDNA for a pea (*Pisum sativum*) seed lipoxygenase was isolated and sequenced. It has a protein coding sequence (2583 bp), 5' (59 bp) and 3' (191 bp) non-coding regions, and a poly(A) tail (20 bp). The predicted amino acid sequence indicates a polypeptide of M_r 97628 that shows about 86% amino acid identity with a soya-bean lipoxygenase 3 protein sequence [Yenofsky, Fine & Liu (1988) Mol. Gen. Genet. 211, 215–222]. The cDNA directs the transcription of mRNA that can be translated to give an anti-lipoxygenase-precipitable polypeptide *in vitro*.

INTRODUCTION

Lipoxygenases (EC 1.13.11.12) are a class of non-haem-iron-containing oxidoreductases whose activity can destroy polyunsaturated fatty acids and produce undesirable tastes and odours during storage of vegetables and seeds [1]. In contrast, lipoxygenase-active soya-bean flour is used in the bread industry to improve dough rheology, lower staling rates and promote bleaching of carotenoid pigments [2]. Legume seeds are a particularly rich source of these enzymes. In *Glycine max* (soya bean), where they constitute about 2% of the total seed protein, polypeptides corresponding to three major isoenzymes (lipoxygenases 1, 2 and 3) have been separated by electrophoresis on SDS/polyacrylamide gels [3,4]. *Pisum sativum* (pea) seeds appear to contain much lower amounts of lipoxygenase 1-like activity [5], and the two major isotypes are assumed to be analogous to lipoxygenases 2 and 3 [6]. By use of an antiserum that was raised against soya-bean lipoxygenase 1, but that showed cross-reactivity to lipoxygenases 2 and 3, two polypeptides were identified on SDS/polyacrylamide gels of *Pisum* seed extracts by Western blotting and a partial cDNA clone corresponding to the larger of these was isolated [7]. Here we report the sequence of a homologous near-full-length cDNA, which can direct lipoxygenase expression *in vitro*.

MATERIALS AND METHODS

The preparation and size fractionation of polyadenylated RNA from developing *Pisum sativum* cultivar Birte and hybrid selection, translation *in vitro* and examination of translation products by SDS/polyacrylamide-gel electrophoresis and fluorography were carried out as previously described [7]. Double-stranded cDNA was synthesized by using an Amersham cDNA kit and blunt-ended cDNA was cloned into the *Sma*I site of plasmid pUC19. The resulting library was screened with insert of plasmid pCD45 [7] that had been 32 P-labelled by nick translation [8]. A near full-length cDNA (pPE1036) was selected and its insert sequenced by standard techniques [9–13]. DNA was sequenced on both strands and across

all restriction sites used. The primer extension method [14] used reverse transcriptase purchased from Bethesda Research Laboratories (GIBCO, Paisley, Strathclyde, U.K.) and transcription *in vitro* was carried out with the Bluescript vector system from Stratagene (Northumbria Biologicals, Cramlington, Northumberland, U.K.) according to the manufacturers' specifications. All other reagents and enzymes were as described previously [7].

RESULTS AND DISCUSSION

Isolation and sequencing of pea seed lipoxygenase cDNA

A small plasmid cDNA library (1300 clones) was constructed from size-selected polyadenylated RNA and screened with a partial-length cDNA (pCD45) insert corresponding to a pea seed lipoxygenase [7]. Of 50 positive clones, four had inserts of approx. 3 kb, the size expected for full-length cDNA [7]. The cDNA sequence of pPE1036 and its deduced amino acid sequence are shown in Fig. 1. The M_r of the polypeptide predicted from this sequence is 97628.

Comparison with soya-bean lipoxygenase sequences

Complete or partial sequences of three soya-bean lipoxygenases have been reported. The cDNA sequence of lipoxygenase 1 [15] has been substantiated by protein sequencing. A partial cDNA (pLX-65) sequence of a 'lipoxygenase 1-like' protein from soya bean [16], after some minor corrections, shows about 84% aligned identity of its amino acid sequence with that of lipoxygenase 1, with 89% identity at the nucleotide level. The coding sequence of soya-bean lipoxygenase 3 has recently been deduced from genomic and cDNA clones [17], revealing 70% and 73% identity respectively between it and those of the lipoxygenase 1 and the 'lipoxygenase 1-like' cDNAs. It has been suggested that the 'lipoxygenase 1-like' cDNA sequence encodes soya-bean lipoxygenase 2 [17].

Alignment of the amino acid sequence predicted by our pea lipoxygenase cDNA with that of soya-bean lipoxygenase 1 reveals approx. 69% amino acid identity,

* To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number X07807.

10 20 30 40 50 60 70 80 90 100 110 120
ATCATCATCATCATCATCATATATATTAGTGTTCCTGTTACTAAGTTGCAAGCAAAGATGTTTTTCAGGCGTGACTGGTATTCTGAATAGAGGCCATAAAATAAAGGGACAGTGGTG
 M F S G V T G I L N R G H K I K G T V V
 130 140 150 160 170 180 190 200 210 220 230 240
 TTGATGAGAAAGATGTGTTGGATATAAACAGCTTAACCACTGTTGGTGGTGTATCGGTCAAGGCTTCGACATTCTCGGTTCCACGGTTGATAATCTCACTGCTTTCTTAGGCCGTTCT
 L M R K N V L D I N S L T T V G G V I G Q G F D I L G S T V D N L T A F L G R S
 250 260 270 280 290 300 310 320 330 340 350 360
 GTTTCCTTCAGTTGATCAGTGCACAAACCTGATGCGACTGGGAAAGGGAAGCTTGGAAAAGCTACTTTCTGGAAGGTATAATTTCTTCATTACCAACTTTGGGAGCTGGACAATCT
 V S L Q L I S A T K P D A T G K G K L G K A T F L E G I I S S L P T L G A G Q S
 370 380 390 400 410 420 430 440 450 460 470 480
 GCTTTTAAGATTCACTTTGAATGGGATGATGATATGGGAATTCCTGGCGCATTTTATATCAAAAACCTTTATGCAAACTGAGTTTTCTTGTGAGTTTGACTCTTGATGACATTCCAAAT
 A F K I H F E W D D D M G I P G A F Y I K N F M Q T E F F L V S L T L D D I P N
 490 500 510 520 530 540 550 560 570 580 590 600
 CATGGAAGTATCTACTTTGTATGCAACTCTTGGATATATAATGCCAAACACCACAAAATCGATCGCATTTTCTTTCGGAATCAGACGTATCTTCGAGTGAGACTCCGGCTCCATTAGTC
 H G S I Y F V C N S W I Y N A K H H K I D R I F F A N Q T Y L P S E T P A P L V
 610 620 630 640 650 660 670 680 690 700 710 720
 CATTATAGAGAAGAATAATTTAAGAGGAGACGGAACAGGAGAGCGCAAAAGATGGGAAAGGATCTATGATTACGATGTCTATAATGATTAGGAAATCCGGATCCGGGTGAG
 H Y R E E E L N N L R G D G T G E R K E W E R I Y D Y D V Y N D L G N P D S G E
 730 740 750 760 770 780 790 800 810 820 830 840
 AACCATGCTCGTCCGGTTCTTGGAGGGTCTGAAACCTATCCTTATCCGCGTAGGGGGAGAACCGGTCGAAACCAACAGAGAAAGATCTAATAGTGAGAGTAGAAGTGATTATGTGTAT
 N H A R P V L G G S E T Y P Y P R R G R T G R K P T R K D P N S E S R S D Y V Y
 850 860 870 880 890 900 910 920 930 940 950 960
 CTTCCAAGAGATGAAGCTTTGGTCACTTGAAGTCATCAGATTTTCTCACTTATGAGTAAAGCTGTGTCTCAAAACGTTGGTCTCTGCAATAGAAATCTGTGTTCTTTGATTGAATTC
 L P R D E A F G H L K S S D F L T Y G L K A V S Q N V V P A L E S V F F D L N F
 970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
 ACGCCAAATGAGTTTGTAGCTTTGATGAAGTTTATGAGTGTATGAGGGTGAAGTAAAGCTGCAACAAATATATCTAGCCAGATAAGCCCGTTACCCGTTGCTTAAGGAGATCTTTTCGA
 T P N E F D S F D E V H G L Y E G G I K L P T N I L S Q I S P L P V L K E I F R
 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 ACCGATGGTGAATAACCTTTAAATATCCACCACCTAAAGTAATCAAGTAAGTAGGTCTGGATGGATGACTGATGAAGAATTCGCAAGAGAAATGCTTGTGGAGTAATCCAAATGTG
 T D G E N T L K Y P P P K V I Q V S R S G W M T D E E F A R E M L A G V N P N V
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
 ATATGCTGTCTTCAGGAGTTTCTCCGCGAAGCAAGCTTGACAGTCAATCTATGGTGATCACACTAGTAAAGCAAGAGACCTAGAGCCTAATTTAGAGGAGCTCAGATAGAA
 I C C L Q E F P P P R S K L D S Q I Y G D H T S K I S K E H L E P N L E G L T V E
 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
 GAGGCAATTCAAACAAAGAAATGTTCTGCTAGATCATCATGACTCAATATGTCCTACTTAAAGGAGATAAACTCAACTCCACAAAGGCTTATGCTACTAGAACCAATCTTTCTTG
 E A I Q N K K L F L L D H H D S I M P Y L R R I N S T S T K A Y A T R T I L F L
 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
 AACAAATACCAAAATTTGAAGCCACTGGCTATAGAGTTGAGTTTACCACACCTCAAGGAGATGAACACGGTGTGTTAGTTATGTTTCAACCTGCGCTAGAGGCGTTGAAAGTTC
 N N N Q N L K P L A I E L S L P H P Q G D E H G A V S Y V Y Q P A L E G V E S S
 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680
 ATTTGGTTATTGCTAAAGCTTATGTGATTGTGAATGATTCTGCTACCATCAACTTGTAGCCATTGGTTGAATACACTCACGAGTTGTTGAGCCGTTTGTGATAGCACTAACAGGCAT
 I W L L A K A Y V I V N D S C Y H Q L V S H W L N T H A V V E P F V I A T N R H
 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 CTGAGTTGCTTACCCGATTATATAAATTTGTATCCTCACTATCGCGATACAATGAATATAAATCTCTCGCTCGACTTTCCCTCGTCAACGACGGCGGTATCATCGAGAAAATTTT
 L S C L H P I Y K L L Y P H Y R D T M N I N S L A R L S L V N D G G I I E K T F
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920
 TTGTGGGAGATATTTCTATGGAATGTCTTAAAGTTTACAAGAAATGGGTTTTTACAGAGCAAGCATTGCTGCTGATCTTATAAAGAGGAATGGCTATTGAGGATCCATCTTCT
 L W G R Y S M E M S S K V Y K N W V F T E Q A L P A D L I K R G M A I E D P S S
 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040
 CCATGTGGTGAAGCTTGGTGGTAGAGGATTCCTTATGCTGTTGATGGAATGAGAGATTTGGGCTATAATAAAGACATGGGTTCAAGACTATGCTCTCTGTACTATACCTCAGATGAA
 P C G V K L V V E D Y P Y A V D G L E I W A I I K T W V Q D Y V S L Y Y T S D E
 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160
 AAAGTTAGGCAAGATTCTGAACCTCAAGCTGGTGAAGAGCTCGTGGAGTCCGTCACGGTGACAGAAACAGAGCCGCTGGTGGCTTAAGATGCAAACTCAGAGAACTTAATTGAA
 K L R Q D S E L Q A W W K E L V E V G H G D K K N E P W W P K M Q T R E D L I E
 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280
 GTTTGAGTATTGTATATGAGTCTTCACTCTTACGCGAGCTGTTAATTTCCGGCAATATTCTACGAGGTTTAACTCTTAACCGCTCAACTCTTAGTAGACGATTATGCTGCTGAG
 V C S I V I W T A S A L H A A V N F G Q Y S Y G G L I L N R P T L S R R F M P E
 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
 AAAGTTGAGTCTGAGTTCGAGGAGCTAGTGAAGTCTCTCAAAAGGCTTACTGAAGACAATTACGCCGAAGTTTCAAGTCTTATTGATCTTTCTGTTATAGAAATATTGTCGAGGCAT
 K G S A E F E E L V K S P Q K A Y L K T I T P K F Q T L I D L S V I E I L S R H
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520
 GCTTCCGATGTTGATCTTGGGAGAGAGACAAATCCGAATTGGACATCGGATAAAGGCTTAGAGGCTTTTAAAGAGTTTGGAAACAGTTGGCTGAAATTTGAGAAGAACTCACT
 A S D E L Y L G E R D N P N W T S D K R A L E A F K K F G N K L A E I E K K L T
 2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640
 CAGAGAAACATGATGAGAAGCTAAGAAATCGTCATGACCGGTTGAGATGCTTATCTTGTCTATCTCTTAGTAAAGAGGACTAATTTTAGAGGAAATCTTAATAGTATCTCC
 Q R N N D E K L R N R H G P V E M P Y T L Y P S S K E G L T F R G I P N S I S
 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760
 ATCTGAAGAGATCTATGTTTCTTAGTTATATATATTGTACCTATGAATAAAGATGCAGAAGAAGGAGTCTATAATCTTCACTACCTGCTCAGTTTGTCTTTGATGCAATTGAGT
 I *
 2770 2780 2790 2800 2810 2820 2830 2840 2850
 GCTTTTCATGTTGAGCATTAAATAATGTATCTGTGTTTGAATGAAGTGTTTTTAAATAATAAATAATCATGTTTGTAAAAAATTTTTTTTTT

Fig. 1. Nucleotide and deduced amino acid sequence of the insert of pPE1036

The sequence underlined represents a 3 bp repeated motif in the 5' untranslated region of the pea cDNA.

Fig. 2. Primer extension analysis of size-fractionated mRNA

A synthetic oligonucleotide primer complementary to nucleotide residues 197–213 of the cDNA was annealed to fractionated mRNA and extended with reverse transcriptase. Lane (a) shows the extended primer fragment sizes examined by electrophoresis on a 6% polyacrylamide sequencing gel with an M13 dideoxy sequence ladder (lanes G, A, T and C) as size markers. The approximate number of bp by which each fragment extends beyond the 5' end of the cDNA insert of pPE1036 is indicated on the left-hand side. A correction of 0.5 bp was applied because of the 5' ^{35}S label of the primer [23]. Lane (b) is a no-RNA control.

Fig. 3. Synthesis of pea seed lipoxygenase polypeptides by the use of transcription and translation systems *in vitro*

Lanes A, B and C represent polypeptides analysed by SDS/polyacrylamide-gel electrophoresis on 15% (w/v) gels. The left-hand arrow indicates the position of the 95000- M_r phosphorylase *a* marker. Lane A, translation products *in vitro* of mRNA hybrid-selected by cDNA clone pPE1036; lane B, translation products *in vitro* of RNA transcribed *in vitro* from the bacteriophage T7 promoter of pSK1036; lane C, anti-lipoxygenase immunoprecipitate of the products corresponding to lane B. Lane D shows the glyoxalated products of transcription *in vitro* of pSK1036 analysed by electrophoresis on a 1.2% (w/v) agarose gel, with the positions of the linearized vector (6 kb) and the transcript RNA (3 kb, arrow) indicated.

with a further 14% similarity if conservative amino acid changes are taken into account. The deduced partial amino acid sequence of pLX-65 shows about 71% identity between it and the corresponding region of the pea sequence. However, there is 86% identity between the aligned predicted amino acid sequences of our cDNA and that corresponding to soya-bean lipoxygenase 3, with a further 7% similarity in respect of conservative amino acid changes. Thus our cDNA clearly encodes a pea seed lipoxygenase polypeptide, the primary structure of which is most similar to soya-bean lipoxygenase 3.

Alignment of the pea lipoxygenase sequence with

those for soya-bean lipoxygenases shows that a region between amino acid residues 31 and 44 of the pea sequence is 71% identical with the corresponding region of lipoxygenase 3. This region is not present in lipoxygenase 1 and highlights the degree of similarity between the primary structure of the pea lipoxygenase and that of soya-bean lipoxygenase 3.

Analysis of the pea lipoxygenase cDNA sequence

Primer extension analysis (Fig. 2) shows three fragments, two of which differ in size by 1 bp and the third of which is relatively minor. We have also cloned and partially sequenced a cDNA that apparently corresponds to the other [7] major pea seed lipoxygenase polypeptide (P. M. Ealing & R. Casey, unpublished work). We know that the oligonucleotide primer used in this experiment should not hybridize to the 5' end of the mRNA corresponding to the latter polypeptide. Thus it appears that there may be one major (within 1 bp) and one minor transcription start site in the gene(s) corresponding to the pPE1036, which results in mRNAs about 18–19 bp and 10 bp longer respectively than the 5' end of our cDNA. Alternatively, the minor fragment may represent premature termination by reverse transcriptase. An unusual feature in the 5' untranslated sequence of the soya-bean lipoxygenase 1 cDNA [15] is a 3 bp repeated motif of TTC. There is a similar repeated motif of ATC in the corresponding region of the pea cDNA (Fig. 1, underlined). The 5' untranslated sequence of the cDNA for soya-bean lipoxygenase 3 does not share this feature and is much shorter [17].

The predicted primary structure of the pea lipoxygenase was analysed by the secondary-structure methods of Chou & Fasman [18]. The α -helix and β -sheet content were 38% and 41% respectively. Hydrophobicity profiles [19] revealed a moderately hydrophobic *N*-terminus but no obvious characteristics of known signal peptide sequences [20]. The hydrophobic region from amino acid residues 709 to 721 contains a large proportion of residues invariant among all lipoxygenase sequences published [21]. There are three hydrophilic regions between amino acid residues 175 and 260, between residues 660 and 690 and between residues 790 and 830. As expected, the predicted secondary structures of soya-bean lipoxygenase 3 and the pea lipoxygenase sequence were very similar. A search of the PIR protein database using the software package FASTP [22] revealed no convincing homologies with any other known proteins.

Transcription and translation from cloned cDNA *in vitro*

As with pCD45, the translation product *in vitro* of mRNA hybrid selected by pPE1036 (Fig. 3, lane A) corresponds in size to the larger of the two major lipoxygenase polypeptides seen on immunoprecipitation of total translation products [7]. We are collaborating with the Carlsberg Laboratories, Copenhagen, in the genetic engineering of strains of baker's yeast to secrete a legume lipoxygenase. We therefore tested the ability of our cDNA to programme the expression of this protein *in vitro* by cloning the cDNA insert of pPE1036 into a transcription vector under the control of the bacteriophage T7 promoter. Fig. 3 (lane D) shows the full-length RNA transcript, which is translated (Fig. 3, lane B) into a polypeptide of apparent M_r approx. 95000 that is immunoprecipitated by anti-lipoxygenase (Fig. 3, lane

C). Thus we are hopeful that our cDNA insert, cloned into the appropriate expression vector, will be capable of directing the expression of pea lipoxygenase in yeast cells.

We thank Dr. Claire Domoney for generously giving plasmid pCD45 and Dr. Suan Cheah for the size-fractionated RNA. This work was supported by the Agricultural and Food Research Council via a grant-in-aid to the John Innes Institute. Support was also received from E.E.C. Contract BAP-0063-UK.

REFERENCES

1. Galliard, T. & Chan, H. W. S. (1980) in *The Biochemistry of Plants* (Conn, E. E. & Stumpf, P. K., eds.), vol. 4, pp. 131–161, Academic Press, London and New York
2. Frazier, P. J. (1979) *Baker's Dig.* **53**, 8–29
3. Kitamura, K., Davies, C. S., Kaizuma, N. & Nielsen, N. C. (1983) *Crop Sci.* **23**, 924–926
4. Kitamura, K. (1984) *Agric. Biol. Chem.* **48**, 2339–2346
5. Yoon, S. & Klein, B. P. (1979) *J. Agric. Food Chem.* **27**, 955–962
6. Reynolds, P. A. & Klein, B. P. (1982) *J. Food Sci.* **47**, 1999–2003
7. Casey, R., Domoney, C. & Nielsen, N. C. (1985) *Biochem. J.* **232**, 79–85
8. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
10. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
11. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
12. Staden, R. (1986) *Nucleic Acids Res.* **14**, 217–231
13. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
14. Brown, J. W. S., Feix, G. & Friendewey, D. (1986) *EMBO J.* **5**, 2749–2758
15. Shibata, D., Steczko, J., Dixon, J. E., Hermodson, M., Yazdanparast, R. & Axelrod, B. J. (1987) *J. Biol. Chem.* **262**, 10080–10085
16. Start, W. G., Ma, Y., Polacco, J. C., Hildebrandt, D. F., Freyer, G. A. & Altschuler, M. (1986) *Plant Mol. Biol.* **7**, 11–23
17. Yenofsky, R. L., Fine, M. & Liu, C. (1988) *Mol. Gen. Genet.* **211**, 215–222
18. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
19. Kyte, M. & Doolittle, F. R. (1982) *J. Mol. Biol.* **157**, 105–132
20. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5166
21. Dixon, R. A. F., Jones, R. E., Diehl, R. E., Bennett, C. D., Kargman, S. & Rouzer, C. A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 416–420
22. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441
23. Donson, J., Morris-Krsinich, B. A. M., Mullineaux, P. M., Boulton, M. I. & Davies, J. W. (1984) *EMBO J.* **3**, 3069–3073

Received 17 March 1988/16 May 1988; accepted 23 May 1988