

Plant Gene Register

Genomic Nucleotide Sequence of a *Brassica napus* 20-Kilodalton Oleosin Gene¹

Keunmyoung Lee and Anthony H. C. Huang*

Department of Botany and Plant Sciences, University of California, Riverside, California 92521

The storage triacylglycerols in seeds are confined to discrete spherical organelles called oil bodies (10). Each oil body of 1 μm diameter contains a matrix of triacylglycerols surrounded by a layer of phospholipids embedded with abundant and unique proteins called oleosins (2, 6). Oleosins are hydrophobic proteins of low M_r , ranging from 16 to 26 kD, depending on the isoforms and plant species. They are present in the seeds of both dicotyledons and monocotyledons. Their possible functions include stabilizing the oil bodies in an aqueous environment and providing signal recognition for the specific binding of lipase during germination. The partial amino acid sequences of the oleosins from maize (16 kD) (9) and *Brassica* (20 kD) (4) and the complete amino acid sequences of the oleosins from maize (18 kD) (5) and carrot (19 kD) (1) have been deduced from their cDNA and genomic sequences, respectively.

Here we report the complete nucleotide sequence of a *Brassica napus* oleosin gene (Fig. 1; Table I). The *Brassica* oleosin gene possesses one intron, whereas all the previously reported oleosin genes (two from maize and one from carrot) do not contain introns. The intron in the *Brassica* oleosin gene occurs between the two exons encoding the central hydrophobic domain and the C-terminal domain, respectively.

The deduced amino acid sequence of the *Brassica* oleosin shows the same three basic structural domains that are common to the four sequenced oleosins. These three domains are an N-terminal hydrophilic domain, a central hydrophobic domain, and a C-terminal amphipathic domain. The *Brassica* oleosin shares significant homology in amino acid sequences at the central hydrophobic domain, but not at the other two domains, with the other reported oleosins. The *Brassica* oleosin does not possess appreciable cleavable signal sequence at the N-terminus, and this is in agreement with its M_r of 20,000 estimated by SDS-PAGE being similar to its M_r of 20,682 derived from the amino acid sequence.

The N-terminal amino acid sequence of the *Brassica* oleosin bears little similarity with those of maize 18-kD oleosin and carrot 19-kD oleosin. This is intriguing, because the maize 18-kD oleosin is correctly targeted to the seed oil bodies in *Brassica* transformed with the maize oleosin gene (3). Presumably, the intracellular targeting signal for oleosin resides in

Table I. Characteristics of a 20-kD Oleosin Gene from *Brassica napus*

Organism:	<i>Brassica napus</i> , L., var Bridger
Location on Chromosome:	Unknown
Function:	Oleosin on the surface of seed oil bodies serving to stabilize the organelles and possibly acting as a receptor for lipase binding during germination.
Gene designation:	BNO20
Source:	Obtained from a λ EMBL-3 genomic library (from Clontech, Palo Alto, CA) using maize 16-kD oleosin cDNA as a probe (9). Fragments subcloned in pUC118 and pUC119 and sequenced by dideoxy sequencing of both strands.
Methods of Identification:	Sequence identical with a segment of the amino acid sequence of an isolated oleosin (8).
Expression and Regulation:	Oleosin synthesized only in maturing seeds. Representing a few percent of the total seed proteins. Oleosin genes in maize (2), carrot (1), and <i>Brassica</i> (7) positively regulated by ABA. Putative ABA regulatory elements present at the 5' upstream (Fig. 1).
Features of gene structure:	An open reading frame of 561 bp ² interrupted by an intron of 462 bp. Intron occurring at the junction of reading frame encoding the central hydrophobic domain and the C-terminal domain. Putative regulatory sequences underlined (Fig. 1).
Codon Usage:	51.3% G + C in the third nucleotide.
(G + C) Content:	51.5% in the coding region.
Structural Features of Protein:	ORF of 561 bp encoding a polypeptide of 187 amino acid residues of M_r 20,682, a value similar to the M_r of the isolated oleosin. Three structural domains present in the protein (N-terminal hydrophilic domain, central hydrophobic domain, and C-terminal amphipathic domain). Belonging to the "low M_r " oleosin isoform.
Antibodies:	Prepared earlier (8).
Subcellular Location:	Localized on the surface of oil bodies.
EMBL Accession No.	M 63985

¹ Support was received from National Science Foundation grant DMB 88-05618.

² Abbreviation: bp, base pair.

-949 CTTTCGGGATAAAGCAATCACCTGGCGATTCAACGTGGTCGGATCATGA
 -900 CGTTCCAGAAAACATCGAGCAAGCTCTCGAAGCAACCAAAGCTGACCTCTTTCCGGATCGTACAGAACCCGAACAATCTCGTTATGTCC
 -810 GTCGTCTCCGAACAGACATCCTCGTAGCTCGGATTATCGACGAATCCATGGCTATACCCAACCTCGTCTTCGTACGCCTGGAACCCCTC
 -720 TGGTACGCCAATTCGGCTCCCCAGAAGCAACCGCGCCGAATTTGGCGGAATTTGCTGACTGGAGACCGGAACATCGTCTCGGGTCCTTG
 -630 GCGATTGGCGGCGGAAGCCGGGTCCGGTGGGGACGAGACCCGAATCCGAGCCTGGTGAAGAGGTTGTTTCATCGGAGATTTATAGACGGGAG
 -540 ATGGATCGAGCGGTTTTGGGAAAGGGGAAGTGGGTTGGCTCTTTTGGATAGAGAGAGTGCAGCTTTGGAGAGAGACTGGAGAGGTTTA
 -450 GAGAGAGACGCGCGGAGATTACCGGAGAGAGCGGAGAGAGATAGCATTATCGAAGGGAAGGGGAAAGAGTGACGTGGAGAAATAGAA
 -360 AACCGTTAAGAGTCCGATATTTATCATATTAAGAAGCCCAATGGGCCTGAACCCATTTAAACAAGACAGATAAATGGGCGGTGTGTTAAGT
 -270 TAACAGAGTGTTAAGCTTCGGTTTCAAATGCCAACCCATAGGAACAAACAAACCTGTCTCAAGTAAACCCCTGCCGTTTACACCTCA
 -180 ATGCGTGCATGTGAAGCCATTAACAACGTGGCGTAGGATGCATGCAGCCATTGACACCTGACTCTCTTCCTTTCTCATATATCTC
 -90 TAATCAATTCAACTACTCATTTGTCTATAGCTATTTCGAAATACATACACATCTTTTCTCTTCGATCTCTCTCAATTACACAAGAACAA
 1 ATGACGGATACAGCTAGAACCATCAGATATCAAGTGCAGATCAGTATGCCCGAGACCGAGACCAGTATTCATGATCGGTCGAGAC
 M T D T A R T H H D I T S R D Q Y P R D R D Q Y S M I G R D
 91 CGAGACAAGTATTCATGATTGGCCGAGCCGAGTACCAAGTCAACATGTATGGTTCGAGACTACTCCAAGTCTAGACAGATGCTAAGGCT
 R D K Y S M I G R D R D Q Y N M Y G R D Y S K S R Q I A K A
 181 GTTACCGCAGTACGGCCGGTGGGTCCCTTCTGTCTCTCCAGTCTCACCTTTGTTCGGAAGTTCATGCTGACTGTTGCGACTCCT
 V T A A V T A G G S L L V L S S L T L V G T V I A L T V A T P
 271 CTGCTGTTATCTTTAGTCCAATCTTGCCCTGCTCATACCCTGGTTCATGCTATCCCGGTTTCTCTCCTCTGGTGGCTTTGGC
 L L V I F S P I L V P A L I T V A L L I T G F L S S G G F G
 361 ATTGCAGCTATAACCGTCTTCTTTGGATCTACAAGtaagtggacattttaaacatataattcaagttgtacaatatgttttaagaagcgg
 I A A I T V F S W I Y K
 451 taattttttttttttttttttttttttttgaatttaagaattcaggggtttccccaaggcttcttaggcccaggactgggtccccctctggcg
 541 ctgacgagctccatgtaataatgccccagtgccgagagaattggttcagcgtgaggttcgaaaccggcgtattgggaagcggtaa
 631 tttattaataaaacataatgggtgaaatagcgacatgccttgtaggggaaaaaaagtacaaaccataaaaattatacataaccgacaa
 721 gtaggattttagatattacattaaaatgcccgttttaccatcatcattttggcctagctatacacacaagacttgacttagcttagctgacgta
 811 cgtgtagtatatgtagcatgcacgtgtgtgtaattgtgatgaatagGTATGCAACGGGAGAGCACCCACAAGGTCAGATAAAGTGGAC
 Y A T G E H P Q G S S D K L D
 901 AGTGCAAGGATGAAGCTGGGAGGCAAAGTTCAGGATATGAAGGACAGAGCTCAGTACATGGACAACAGCAAAGGTTGGGAAACGAC
 S A R M K L G G K V Q D M K D R A Q Y Y G Q Q Q T G G E H D
 991 CGTGACCGTACCCGTTGAAACCCAGCACACTACCTAAATTCAGCCATGACTATTTTCATAGTCCAATAAGGCTGATGTGGGAGTCCAGTT
 R D R T R G T Q H T T *
 1081 TATGAGCAATAAGGTTTATAGAAATTGATCAATGTTTATAATAAAGGGGGAAGATGATATCACAGTCTTTTCTTTTGGCTTTTGT
 1171 TAAATTTGTGTGTTTCTATTTGTAACCTTCTGTATATGTTGTAATTTCTTCCCTTTTAAAGTGGTATCGTCTATATGGTAAAACGTTAT
 1261 GATTGGTCTTTCCCTTTCTGTTTATAGGATAAAAGACTGCATGTTTATCTTTAGTTATATTGTTGAGTAAATGAACCTTTATAGAT
 1351 CTGTTCCGTAGAGTAGACTAGCAGCCGAGTTGAGCTGACTGAACCTGCTGGCAATGTGAACACTGGATGCAAGATCAGATGTGAAGATC
 1441 TCTAATATGGTGGGATTGAACATATCGTGTCTATATTTTGGTGGCATTAAAGCTCTTAACATAGATTAACACTGATGCGAGTCAATGGT
 1531 TCATACACATATATAGTAAGGAATTACAATGGCAACCCAACTTCAAAAACAGTAGGCCACCTGAATTGCCCTTATCGAATAAGAGTTGT
 1621 TTCCCCCTTTTTCATGGATGTAATACATGGGATTGGGAGTTTGAATGAACGTTGAGACATGGCAGAACC

Figure 1. A 2640-bp² sequence of a *B. napus* 20-kD oleosin gene. The transcription initiation codon ATG is numbered as the first three nucleotides. The putative regulatory sequences are underlined, including TATA box, CAAT boxes, RY repeats, GC boxes, octamer, putative ABA-binding sites (ACGTGGCGT and ACGTGTCTC), and polyadenylation signal. The intron sequence is printed in lower case letters. The predicted amino acid sequence is shown. The amino acid sequence, which was determined earlier by peptide microsequencing (8), is underlined.

the highly conserved central hydrophobic domain or in the yet-unidentified secondary structures in the N- or C-terminus.

Earlier, we identified at least two isoforms of oleosins in seeds from different, and within the same, species (8). The two isoforms are distinguishable immunologically, and in general, one isoform has a M_r higher than that of the other isoform within the same species. The *Brassica* 20-kD oleosin and the maize 16-kD oleosin apparently belong to the same "low M_r " isoform, whereas the carrot 19-kD and maize 18-kD represent the "high M_r " isoform. This category is revealed in a comparison of the amino acid sequences among the oleosins: 48% identical amino acid residues between the two high M_r oleosins (maize 18-kD oleosin and carrot 19-kD oleosin), 51% between the two low M_r oleosins (maize 16-kD oleosin and rapeseed 20-kD oleosin), 37% between the two maize oleosins, and 33% or less between any other pair of high and low M_r oleosins. The findings that each pair of oleosin isoforms are more related between themselves than to the other oleosin isoforms explains our earlier observation of their immunological kinship (8).

LITERATURE CITED

- Hatzopoulos P, Franz G, Choy L, Sung RZ (1990) Interaction of nuclear factors with upstream sequences of a lipid body membrane protein gene from carrot. *Plant Cell* 2: 457-467
- Huang AHC, Qu R, Lai YK, Ratnayake C, Chan KL, Kuroki GW, Oo KC, Cao YZ (1991) Structure, synthesis and degradation of oil bodies in maize. *In* MJ Emes, ed, *Compartmentation of Plant Metabolism in Non-Photosynthetic Tissues*. Cambridge University Press, Cambridge, England, pp 43-58
- Lee WS, Tzen JTC, Kridl JC, Radke SE, Huang AHC (1991) Maize oleosin is correctly targeted to seed oil bodies in *Brassica napus* transformed with the maize oleosin gene. *Proc Natl Acad Sci USA* in press
- Murphy DJ, Keen JN, O'Sullivan JN, Au DMY, Edwards EW, Jackson PJ, Cummins I, Gibbons T, Shaw CH, Ryan AJ (1991) A class of amphipathic proteins associated with lipid storage bodies in plants. Possible similarities with animal serum apolipoproteins. *Biochim Biophys Acta* 1088: 86-94
- Qu R, Huang AHC (1990) Oleosin KD18 on the surface of oil bodies in maize: genomic and cDNA sequences, and the deduced protein structure. *J Biol Chem* 265: 2238-2243

6. **Qu R, Wang SM, Lin YH, Vance VB, Huang AHC** (1986) Characteristics and biosynthesis of membrane proteins of lipid bodies in the scutella of maize (*Zea mays L.*). *Biochem J* **234**: 57–65
7. **Taylor DC, Weber N, Underhill EW, Pomeroy MK, Keller WA, Scowcroft WR, Wilen RW, Moloney MM, Holbrook LA** (1990) Storage protein regulation and lipid accumulation in microspore embryos of *Brassica napus L.* *Planta* **181**: 18–26
8. **Tzen JTC, Lai YK, Chan KL, Huang AHC** (1990) Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. *Plant Physiol* **94**: 1282–1289
9. **Vance VB, Huang AHC** (1987) The major protein from lipid bodies of maize. Characteristics and structure based on cDNA cloning. *J Biol Chem* **262**: 11275–11279
10. **Yatsu LY, Jacks TJ** (1972) Spherosome membranes. Half unit-membrane. *Plant Physiol* **49**: 937–943