*Plant Cell Physiol.* 40(10): 1079–1086 (1999)


# Cloning and Secondary Structure Analysis of Caleosin, a Unique Calcium-Binding Protein in Oil Bodies of Plant Seeds

**Jeff C.F. Chen, Corinne C.Y. Tsai and Jason T.C. Tzen** [1]

*Graduate Institute of Agricultural Biotechnology, National Chung-Hsing University, Taichung, Taiwan*

**Plant seed oil bodies comprise a matrix of triacylglycerols surrounded by a monolayer of phospholipids embedded with abundant oleosins and some minor proteins. Three minor proteins, temporarily termed Sops 1–3, have been identified in sesame oil bodies. A cDNA sequence of Sop1 was obtained by PCR cloning using degenerate primers derived from two partial amino acid sequences, and subsequently confirmed via immunological recognition of its over-expressed protein in *Escherichia coli*. Alignment with four published homologous sequences suggests Sop1 as a putative calcium-binding protein. Immunological cross-recognition implies that this protein, tentatively named caleosin, exists in diverse seed oil bodies. Caleosin migrated faster in SDS-PAGE when incubated with $Ca^{2+}$. A single copy of caleosin gene was found in sesame genome based on Southern hybridization. Northern hybridization revealed that both caleosin and oleosin genes were concurrently transcribed in maturing seeds where oil bodies are actively assembled. Hydropathy plot and secondary structure analysis suggest that caleosin comprises three structural domains, i.e., an N-terminal hydrophilic calcium-binding domain, a central hydrophobic anchoring domain, and a C-terminal hydrophilic phosphorylation domain. Compared with oleosin, a conserved proline knot-like motif is located in the central hydrophobic domain of caleosin and assumed to involve in protein assembly onto oil bodies.**

**Key words:** Caleosin — Oil body — Proline knot — Seed — Sesame.

Vegetable cooking oils are triacylglycerols (TAGs) extracted from various plant seeds. Plant seeds store TAGs as food sources for germination and postgerminative growth of seedlings. The storage TAGs are confined to discrete spherical organelles called oil bodies, lipid bodies, oleosomes or spherosomes (Yatsu and Jacks 1972, Murphy 1993, Huang 1996). Oil bodies maintain as individual small organelles even after a long period of storage in plant seeds (Slack et al. 1980). This stability is a consequence of the steric hindrance and electronegative repulsion provided by proteins, mostly a structural protein termed oleosin, on the surface of oil bodies (Tzen et al. 1992). It has been suggested that the entire surface of an oil body is covered by proteins (Tzen and Huang 1992). Therefore, the compressed oil bodies in the cells of a mature seed never coalesce or aggregate.

An oil body, 0.5 to 2.5 μm in diameter (Tzen et al. 1993), contains a TAG matrix surrounded by a monolayer of phospholipids (PLs) embedded with abundant oleosins and probably some minor proteins of higher molecular mass (Tzen et al. 1997). Oleosins are alkaline proteins of molecular mass 15 to 24 kDa depending on species (Qu et al. 1986), and have been extensively investigated in the past decade (Napier et al. 1996). Recently, three minor proteins, temporarily termed Sops 1–3, have been identified exclusively present in sesame oil bodies via immunofluorescence labeling (Chen et al. 1998). However, the biological functions of these three minor proteins remain unknown. Neither is known for any of the genes encoding unique oil-body proteins other than oleosins.

In this study, we cloned a cDNA encoding one of the unique oil-body proteins, Sop1, from maturing sesame seeds. The deduced protein comprises a putative calcium-binding motif, and thus tentatively named caleosin. Immunological detection and sequence analyses suggest that caleosin is a unique protein in diverse seed oil bodies.

## Materials and Methods

*Plant materials*—Mature and fresh maturing sesame (*Sesamum indicum* L.) seeds were gifts from the Crop Improvement Department, Tainan District Agricultural Improvement Station. Mature seeds of sunflower (*Helianthus annus* L.), soybean (*Glycine max* L.), peanut (*Arachis hypogaea* L.), and rape (*Brassica campestris* L.) were purchased from local seed stores. The seeds were soaked in water for 10 min prior to purification of oil bodies.

*Purification of oil bodies*—Oil bodies were extracted from diverse seeds and subjected to further purification using the protocol developed by Tzen et al. (1997) including two-layer flotation by centrifugation, detergent washing, ionic elution, treatment of chaotropic agent, and integrity testing with hexane.

*Elution of caleosin from SDS-PAGE*—Proteins extracted from seed oil bodies were resolved by SDS-PAGE using 12.5% polyacrylamide (Laemmli 1970). After electrophoresis, the gel was

Abbreviations: DEPC, diethyl pyrocarbonate; EGTA, ethylene-glycol-bis(β-aminoethyl ether)-tetraacetic acid; PL(s), phospholipid(s); TAG(s), triacylglycerol(s).

The nucleotide sequence reported in this paper has been submitted to the GenBank Data Bank under accession number AF109921.

[1] To whom correspondence should be addressed.

stained with Coomassie Blue R-250 and destained. After Coomassie Blue staining, protein band corresponding to caleosin was cut and homogenized in extraction buffer (0.125 M Tris-HCl, pH 8.0, 1 mM EDTA, 0.1% SDS) using a pestle and mortar (Chuang et al. 1996). The homogenate was centrifuged at 10,000 × *g* for 10 min. The supernatant was retained, and the pellet was re-extracted with extraction buffer. The re-extraction was performed two more times. The eluted caleosin in the combined supernatants was precipitated with an equal amount of acetone pre-chilled at −20°C. The acetone mixture was kept at −20°C for 30 min and then centrifuged at 10,000 × *g* for 30 min to obtain a pellet. The protein pellet was dissolved in the extraction buffer to a concentration of 1 mg $ml^{-1}$.

*Partial amino acid sequencing*—Caleosin eluted from the SDS-PAGE gel was subjected to trypsin or chymotrypsin digestion. In the reaction mixture, 20 μg caleosin was digested with 5 μg trypsin (bovine pancreas type III, Sigma, St. Louis, MO, U.S.A.) or chymotrypsin (bovine pancreas type II, Sigma, St. Louis, MO, U.S.A.) at 37°C for 30 min in 50 mM Tris-HCl, pH 7.5. After digestion, the reaction mixture was added to an equal volume of 2× SDS-PAGE sample buffer and boiled for 5 min. The hydrolysis products were resolved in a SDS-PAGE gel using 15% polyacrylamide. After electrophoresis, fragments of polypeptide were transferred onto a piece of PVDF membrane (Immobilon-P transfer membrane from Millipore, Bedford, MA, U.S.A.) at a current of 0.5 A for 30 min at 4°C in a blotting buffer of 10% methanol and 10 mM CAPS (3-cyclohexylamino-1-propanesulfonic acid)-NaOH, pH 11. After blotting, the PVDF membrane was stained with Coomassie Blue for 5 min, destained for 5 min, rinsed with water three times, and then left to dry in the air. The major stained band in each digestion was picked up for sequencing from the N-terminus by the Applied Biosystems 476A Protein Sequencer in Chung-Hsing University, Taiwan.

*Isolation of total RNA and poly(A)$^+$ RNA*—Total RNA was extracted from the maturing seeds (24 d after flowering) ground in liquid nitrogen using the phenol/SDS method (Wilkins and Smart 1996). Poly(A)$^+$ RNA was isolated with Dynabeads (Dynal) following the manufacturer's instructions. The isolated poly(A)$^+$ RNA was dissolved in diethyl pyrocarbonate (DEPC)-treated water and then quantitated as $A_{260}$.

*cDNA library construction, PCR cloning, and sequencing*—cDNA was synthesized from poly(A)$^+$ RNA according to the protocol described in the manufacturer's instructions (cDNA synthesis, ZAP-cDNA synthesis, and ZAP-cDNA Gigapack III Gold Cloning kits purchased from Stratagene, La Jolla, CA, U.S.A.). A cDNA library of approximately $10^6$ plaques was constructed with 5 μg poly(A)$^+$ RNA. The plaques were subjected to in vivo excision of the pBluescript phagemid from the Uni-ZAP XR vector following the manufacturer's instructions. Two pairs of degenerate primers (5′-GNGGNYTNGTNGCNCCNGAYAT-G-3′ and 5′-GTCATNACCCANARYTCNCCNARNGT-3′) were designed for both directions according to the first nine residues of the two partial amino acid sequences of Sop1 and PCR amplification was carried out using the excised phagemids as templates. A PCR fragment of approximately 400 bp was harvested, ligated into the pGEM-T Easy Vector systems (Promega, Madison, WI, U.S.A.), and subjected to sequencing using the Sequence™ Version 2.0 DNA Sequence kit (USB). The upstream and downstream sequences of the clone were separately obtained using the same strategy of PCR amplification (except for the two primers in reactions) with a 24-nucleotide primer (5′-GGTTTCA-ATGTGATAGCTTCCCTT-3′ or 5′- TTGAAACCAATTTGGC-GAAGTCCA-3′) from the middle sequence of the obtained PCR fragment and T7 or T3 primer from the phagemid vector. The entire Sop1 clone of 1,030 bp was linked by PCR and sequenced from both directions.

*Over-expression of the sesame caleosin clone in Escherichia coli*—A full-length cDNA clone was constructed in the non-fusion expression vector, pET29a(+) (Novagen, Madison, WI, U.S.A.), using an *Nde*I site at the initial methionine position and a *Xho*I site in the polylinker of the vector. The recombinant plasmid was used to transform *E. coli* strain NovaBlue (DE3). Over-expression was induced by 0.1 mM IPTG in a bacteriophage T7 RNA polymerase/promoter system. Three h after induction, the *E. coli* cells were harvested, crashed by sonication in 10 mM sodium phosphate buffer, pH 7.5, and then subjected to further analyses by SDS-PAGE and immunoassay.

*Antibody preparation and western blotting*—Antibodies against sesame caleosin were raised in chickens and purified from egg yolks for immunoassay (Chen et al. 1998). In the immunoassay, proteins in an SDS-PAGE gel were transferred onto nitrocellulose membrane using a Bio-Rad Trans-Blot system (Bio-Rad, Hercules, CA, U.S.A.) according to the manufacturer's instructions. The membrane was subjected to immunodetection using secondary antibodies conjugated with horse radish peroxidase, and then incubated with 3 mM 4-chloro-1-naphthol containing 0.015% $H_2O_2$ for color development.

*Effects of EGTA and various cations on migration of seed-purified and over-expressed caleosin in SDS-PAGE*—Oil bodies (50 mg) in 500 μl of a 10 mM Tris-HCl buffer, pH 7.5 were mixed with an equal volume of 200 mM ethylene-glycol-bis(β-aminoethyl ether)-tetraacetic acid (EGTA) for 5 min at room temperature. The EGTA-treated oil bodies were collected by centrifugation (10 min at maximum speed in an eppendorf centrifuge) and resuspended in 500 μl of the same Tris buffer. The resuspended oil bodies were incubated with various concentrations of $CaCl_2$, $MgCl_2$, KCl, or NaCl for 5 min at room temperature, followed by 100 mM EGTA treatment. Effects on the migration of proteins extracted from oil bodies of the above diverse treatments were analyzed in an SDS-PAGE gel. Similar effects were observed using over-expressed caleosin in *E. coli* instead of purified oil bodies.

*Isolation of genomic DNA and Southern blot analysis*—Genomic DNA was isolated from sesame leaves according to the protocol described by Sambrook et al. (1989). Isolated genomic DNA of 2.5 μg was digested with *Eco*RI or *Kpn*I at 37°C overnight, and the resulting fragments were resolved in a 0.8% agarose gel and transferred onto a piece of blotting membrane (Sartorius, Göttingen, Germany). The blotted membrane was UV cross-linked and hybridized with a $^{32}$P-labeled probe containing the coding sequence of sesame caleosin at 65°C for 14 h. The membrane was washed four times at 65°C following the manufacturer's instructions and exposed to an X-ray film.

*Isolation of RNA and Northern blot analysis*—Total RNA from various stages of maturing seeds was extracted in liquid nitrogen using the Ultraspec™ RNA isolation system (Biotecx, Houston, TX, U.S.A.). The isolated RNA was dissolved in DEPC-treated water and then quantitated as $A_{260}$. Total RNA of 20 μg was resolved in a 1.2% formaldehyde/agarose gel, transferred onto a piece of blotting membrane (Sartorius, Göttingen, Germany), and then subjected to Northern hybridization using the same probe described in Southern hybridization or a $^{32}$P-labeled probe containing the coding sequence of sesame 15.5 kDa oleosin (Chen et al. 1997).

*Sequence analyses*—Sequence comparisons were performed with the GenBank using Blast program (Altschul et al. 1990). Amphipathic α-helix was predicted using helix wheel projection

(Shiffer and Edmundson 1967) and helical hydrophobic moment (Eisenberg et al. 1982). Phosphorylation sites were identified using Motif program at the GenomeNet, Japan (www.motif.genome.ad.jp). Hydropathy profile was plotted with a window size of 15 using hydropathy index described by Kyte and Doolittle (1982). Protein secondary structure was predicted using amino acid hydrophobicity and several prediction programs in the Proteomics Tools at ExPASy Molecular Biology Server (expasy.hcuge.ch/www/tools.html).

## Results

*Cloning of a cDNA encoding Sop1, a unique protein in oil bodies of sesame seed*—To clone the corresponding gene of sesame Sop1, two partial amino acid sequences, MARGLVAPDMDHPNG and TMPDRLTLGELWSM, were obtained from fragments of purified Sop1 protein cleaved by trypsin and chymotrypsin, respectively. With the relative location of these two partial sequences in Sop1 protein uncertain, two pairs of degenerate primers were designed for both directions to clone its corresponding cDNA by PCR. A full-length cDNA clone (accession no. AF109921) was obtained and sequenced. The cDNA fragment comprises 1030 nucleotides, consisting of a 39-nucle-

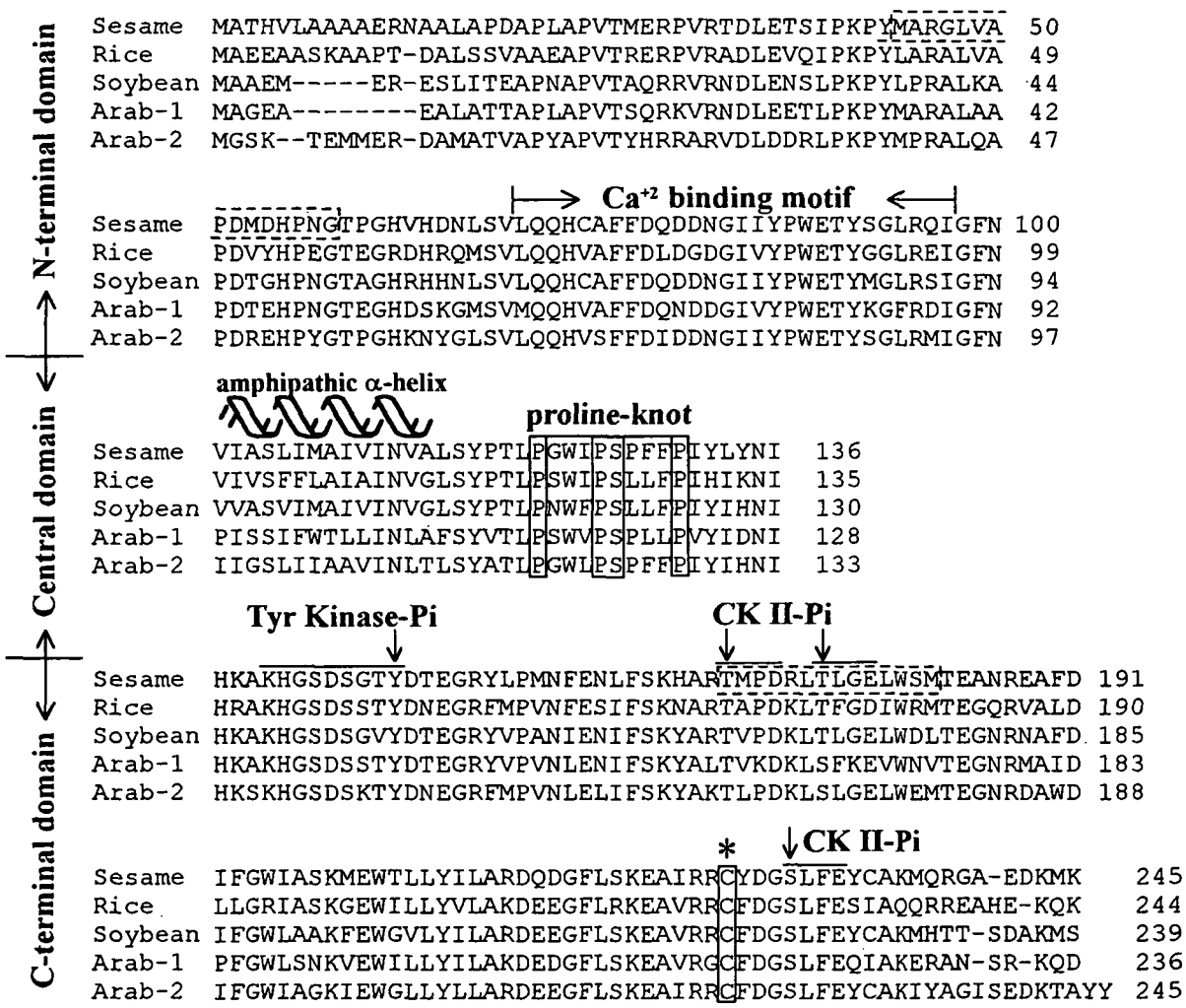


**Fig. 1** Sequence alignment of sesame caleosin with four other available homologous sequences. The sequences are aligned according to the three structural domains (N-terminal, central hydrophobic, and C-terminal domains) of caleosin. The amino acid number for the last residue in each row is listed on the right for each species. Broken lines in the sequences represent gaps introduced for best alignment. The positions of a calcium-binding motif, an amphipathic α-helix, a proline knot-like motif, and four phosphorylation sites (one tyrosine kinase and three casein kinase II phosphorylation sites) are indicated on tops of the sequences. The four invariable residues (three proline and one serine residues) in the proline knot-like motif are boxed. The consensus sequences of the phosphorylation sites are marked by upper lines with the potential phosphorylated residue pointed by arrows. A unique conserved cysteine residue prior to the last phosphorylation site is boxed and indicated by a star. Two partial sequences obtained directly from amino acid sequencing are enclosed with dotted lines. The accession numbers of the aligned sequences are: rice, X89891; soybean, AF004809; Arab-1, AC002332; Arab-2, AF067857.

otide 5′ untranslated region, an open reading frame of 738 nucleotides, and a 253-nucleotide 3′ untranslated region. The open reading frame is deduced to encode a putative calcium-binding protein homologous to four published sequences, and thus tentatively named caleosin (Fig. 1). The deduced sesame caleosin comprises 245 amino acid residues with a molecular mass 27,638 Da.

*Immunological recognition of over-expressed caleosin in E. coli by antibodies against purified Sop1*—The cDNA clone of caleosin was constructed in a non-fusion vector and then over-expressed in *E. coli*. The over-expressed caleosin in *E. coli* was resolved by SDS-PAGE and further detected by immunoassaying using antibodies raised against Sop1 purified from sesame seed oil bodies (Fig. 2). The expressed caleosin and the seed-purified Sop1 migrated to the same distance in SDS-PAGE and equivalently recognized by antibodies against Sop1. The result indicates that caleosin is Sop1, a unique seed oil body protein, and that neither removable signal sequence exists nor post-translational cleavage occurs in mature caleosin.

*Immunological cross-recognition of caleosin in oil bodies of diverse oily seeds*—Proteins extracted from oil bodies of sesame and four other oily seeds (sunflower, soybean, peanut, and rapeseed) were resolved in SDS-PAGE and subjected to immunodetection using antibodies against sesame caleosin (Fig. 3). Polypeptides of molecular masses similar to sesame caleosin were cross-recognized in sunflower, soybean, and rapeseed while a polypeptide of lower molecular mass was detected in peanut. It is likely that caleosin occurs in seed oil bodies of diverse species.

*Caleosin migrating faster in SDS-PAGE when associated with $Ca^{2+}$*—To examine whether caleosin binds to calcium, purified oil bodies and over-expressed caleosin in *E. coli* were separately treated with EGTA to remove endogenous cations and then incubated with various cations. In both seed-purified oil bodies (Fig. 4A) and over-expressed caleosin in *E. coli* (Fig. 4B), caleosin migrated slower after EGTA treatment. This retarded migration could be rescued by adding $Ca^{2+}$, but not three other examined cations, to the cation-depleted (EGTA treated) caleosin. Accordingly, the faster migration of the $Ca^{2+}$-

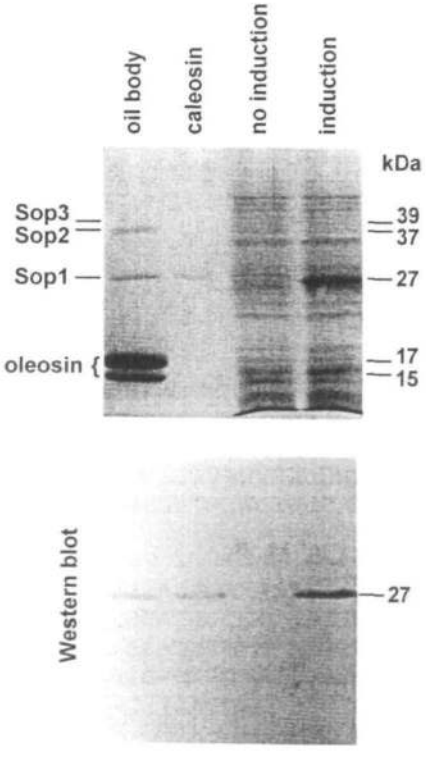


**Fig. 2** SDS-PAGE and Western blotting of the over-expressed sesame caleosin in *E. coli*. Along with sesame oil body proteins (10 μg) and purified caleosin (1 μg), the total proteins (10 μg) of *E. coli* over-expressed in a non-fusion vector before or after IPTG induction were resolved in SDS-PAGE. A duplicate gel was transferred onto nitrocellulose membrane and then subjected to immunoassaying using antibodies against the seed-purified Sop1 protein. Labels on the right indicate the molecular masses of proteins.

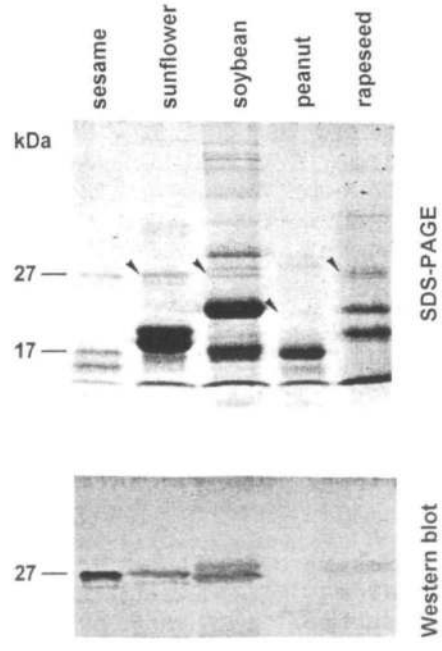


**Fig. 3** SDS-PAGE and Western blotting of proteins extracted from seed oil bodies of various species. Proteins extracted from oil bodies of sesame (4 μg protein), sunflower (20 μg protein), soybean (20 μg protein), peanut (10 μg protein), and rapeseed (10 μg protein) were resolved in SDS-PAGE. A duplicate SDS-PAGE gel (except the sesame lane in which one fifth of protein content was loaded to avoid the effect of over-reaction to the original antigen) was transferred onto nitrocellulose membrane and then subjected to immunoassaying using antibodies against sesame caleosin. The cross-recognized protein bands in the SDS-PAGE gel are pointed by arrows. Labels on the left indicate the molecular masses of proteins.

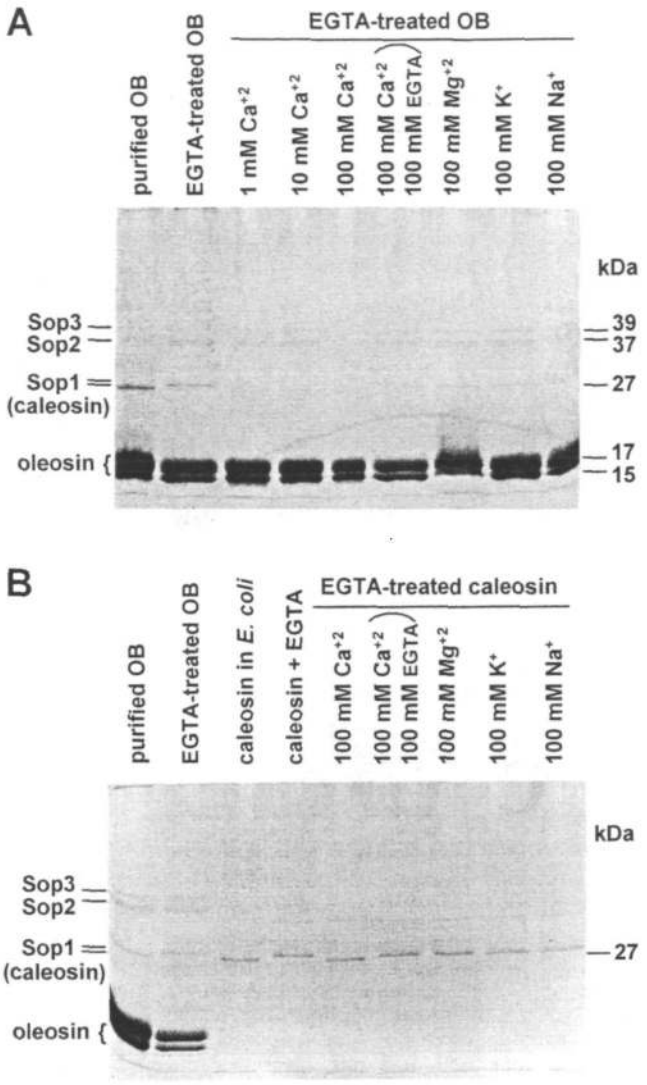


**Fig. 4** SDS-PAGE of sesame oil body proteins (A) and over-expressed caleosin in *E. coli* (B) treated with EGTA and various cations. (A) Oil bodies containing 10 μg protein were treated with EGTA and then incubated with diverse concentrations of $Ca^{2+}$, $Mg^{2+}$, $K^+$, or $Na^+$. The oil bodies incubated with 100 mM $Ca^{2+}$ were further treated with 100 mM EGTA. (B) Caleosin over-expressed in *E. coli* (20 μg protein) was treated with EGTA and then incubated with 100 mM $Ca^{2+}$, $Mg^{2+}$, $K^+$, or $Na^+$. The over-expressed caleosin incubated with 100 mM $Ca^{2+}$ was further treated with 100 mM EGTA. Labels on the right indicate the molecular masses of proteins.

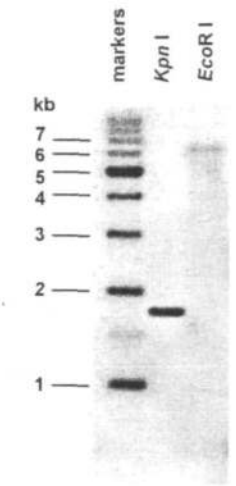


**Fig. 5** Southern blot analysis of genomic DNA extracted from leaves of sesame plants. Each lane was loaded with 2.5 μg of genomic DNA completely digested with *Eco*RI or *Kpn*I. After blotting, the membrane was hybridized with a $^{32}P$-labeled probe containing the coding sequence of sesame caleosin.

*A single copy of caleosin gene specifically transcribed during seed maturation*—To detect the copy number of caleosin gene in sesame genome, the cDNA clone of caleosin was $^{32}P$-labeled as a probe to hybridize genomic DNA digested with two restriction enzymes (Fig. 5). One single fragment of approximately 6–7 kb or 1–2 kb was detected after *Eco*RI or *Kpn*I digestion. It is assumed that a single copy of caleosin gene exists in sesame genome. To

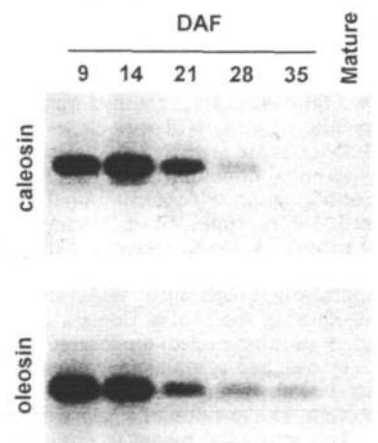


**Fig. 6** Northern blot analysis of total RNA extracted from various stages of maturing sesame seeds. Each lane was loaded with 20 μg of total RNA extracted from maturing seeds at various days after flowering (DAF). After blotting, the membrane was hybridized with a $^{32}P$-labeled probe containing the coding sequence of sesame caleosin (upper panel) or 15.5 kDa oleosin (lower panel). Only the portion of the membrane corresponding to the visible hybridized RNA is shown.

associated caleosin could be retarded by EGTA treatment. No migration shift was observed for other oil body proteins (Sop2, Sop3, and oleosins) in the above treatments. The results propose that caleosin is a calcium-binding protein, and that the $Ca^{2+}$-associated caleosin renders a more compact structure and migrates faster than its $Ca^{2+}$-free conformation.

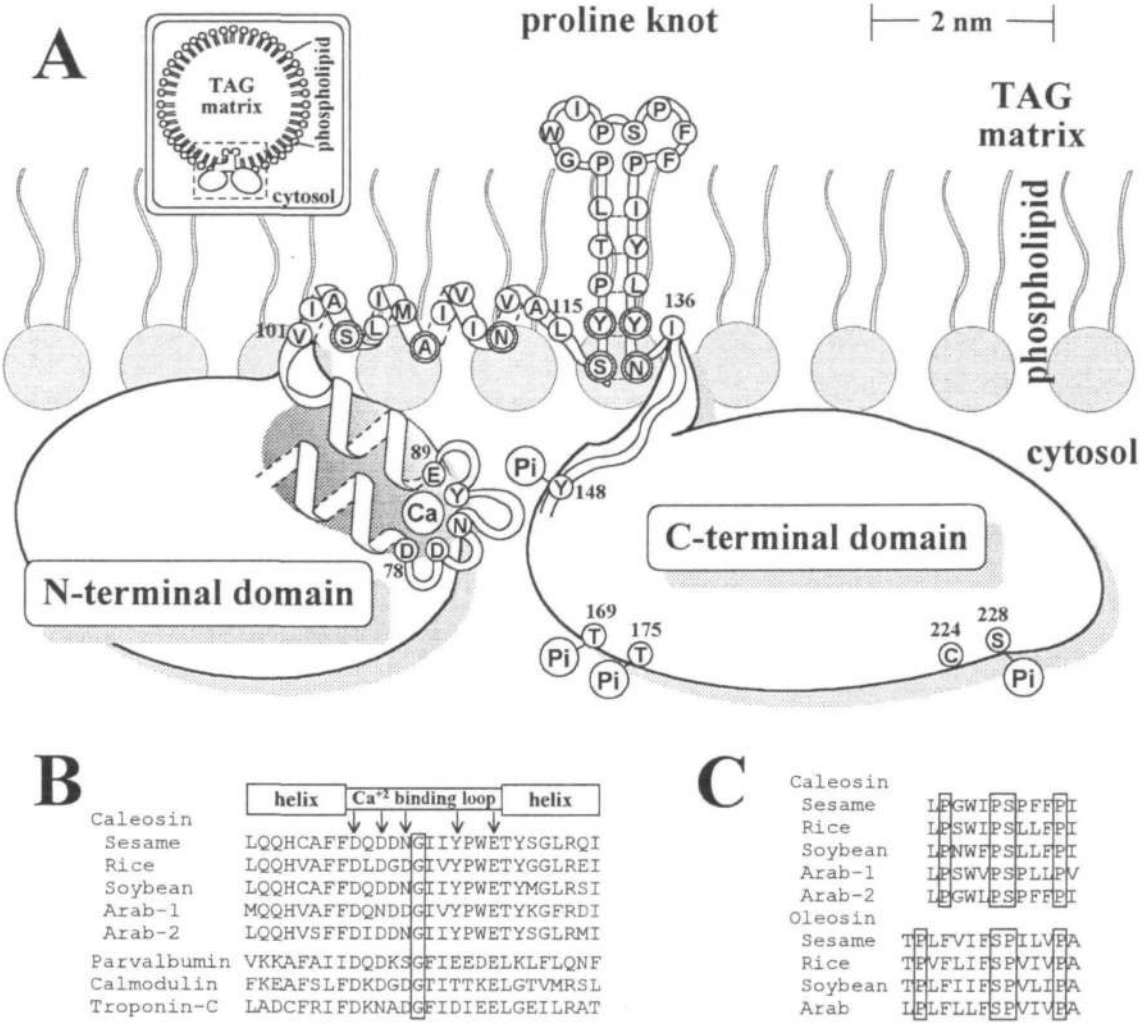


**Fig. 7** (A) A secondary structural model of sesame caleosin on the surface of an oil body. A monolayer of PLs, depicted by gray circles attached with two tails, segregates the hydrophobic TAG matrix of an oil body from hydrophilic cytosol. Amino acid residues are represented by one-letter symbols in open circles. Numbers next to residues represent their relative positions counting from N-terminus. Three structural domains are predicted in a caleosin molecule: N-terminal hydrophilic, central hydrophobic, and C-terminal hydrophilic domains. The N-terminal domain (residues 1–100) is exposed to the cytosol and contains a calcium-binding motif (see [B] for details) preceding the central doamin. Five calcium-binding ligands (oxygen-containing residues) are shown around a calcium atom. The central domain (residues 101–136) composed of an amphipathic *a*-helix and a proline knot-like region is mostly embedded in the PL layer, and thus anchors the protein on the surface of the oil body. The proline knot-like region comprises a similar structure found in the proline knot motif of oleosin (see [C] for details). Two short anti-parallel $\beta$-strands connected to the proline knot are paired with amino acid residues of comparable hydrophobicity and stabilized by forming hydrogen bonds as shown in dashed lines linking residues of the two strands. Most residues in the central domain are hydrophobic and surrounded by hydrophobic tails of PLs while seven relatively hydrophilic residues (double circled) are located in the hydrophilic area of PL head groups. The C-terminal domain (residues 137–245) is exposed to the cytosol and contains four phosphorylation sites (one tyrosine kinase and three casein kinase II phosphorylation sites) conserved among the known sequences. An invariable cysteine residue prior to the last casein kinase II phsophorylation site occurs near the end of the protein. The location of a caleosin molecule (enclosed by dashed lines) in an entire oil body is shown in the inserted panel on the left top. This structural model is applicable to all the other four homologous sequences of caleosin in various species. (B) Sequence comparison among calcium-binding motifs of caleosin and three other calcium-binding proteins. The predicted two helices and one calcium-binding loop in an EF hand motif are indicated on top of the aligned sequences. Based on the alignment, the essential glycine residue acting as a structural turning point is boxed while the five conserved oxygen-containing residues serving as calcium-binding ligands are indicated by arrows. The accession numbers of sequences encoding the above calcium-binding proteins are: parvalbumin, 225026; calmodulin, 115508; troponin-C, 231093. (C) Sequence comparison between the proline knot-like motif of caleosin and the proline knot motif of oleosin. Four conserved residues (three proline and one serine residues) in both caleosin and oleosin are boxed. The unique serine residue adjacent to a conserved proline residue occurs around the center of the sequences whereas in a reverse order in these two comparable motifs. The accession numbers of the four oleosin sequences are: sesame, U97700; rice, AF019212; soybean, V09119; Arabidopsis (Arab), X62353. The conserved proline knot motif shown in these four species exists in all known oleosins from diverse species.

inspect transcription of caleosin gene, mRNAs in various stages of maturing seeds were extracted and subjected to Northern hybridization using the same probe described in the above Southern hybridization (Fig. 6). Accumulation of caleosin mRNA peaked in maturing seeds approximately two weeks after flowering, diminished thereafter, and vanished in mature seeds in a mode similar to that of oleosin mRNA. The result reveals that caleosin gene along with oleosin gene is transcribed during seed maturation when oil bodies are actively assembled. This observation is in accord with the exclusive accumulation of caleosin and oleosin proteins in oil bodies of maturing sesame seeds detected by Western blots (Chen et al. 1998).

*Three structural domains in caleosin predicted by sequence analyses*—Similar to oleosin, caleosin possesses a long stretch of hydrophobic residues in the center region of the protein according to the hydropathy plot (data not shown). In a more detailed analysis, caleosin comprises three structural domains including an N-terminal hydrophilic domain, a central hydrophobic domain, and a C-terminal hydrophilic domain (Fig. 7A). In the N-terminal hydrophilic domain, a helix-turn-helix calcium-binding motif occurs prior to the central hydrophobic domain. This calcium-binding motif consists of an EF hand sequence of 28 residues including an invariable glycine residue as a structural turning point and five conserved oxygen-containing residues as calcium-binding ligands (Fig. 7B).

The central hydrophobic domain can be divided into an amphipathic $\alpha$-helix and an anchoring region. The amphipathic $\alpha$-helix is composed of more hydrophobic residues than hydrophilic ones, and thus probably embedded deeply in the surface area of an oil body which provides an interface between hydrophobic and hydrophilic environments. The anchoring region comprises a proline knot-like motif and a pair of anti-parallel $\beta$-strands. The proline knot-like motif contains four invariable residues (three proline and one serine residues) surrounded by mostly hydrophobic residues in a similar organization found in oleosin (Fig. 7C). The two anti-parallel $\beta$-strands are stabilized by interstrand hydrogen bonds and paired with amino acid residues of comparable hydrophobicity (Fig. 7A) in a similar manner as appeared in the paired anti-parallel $\beta$-strands of oleosin (Tzen et al. 1992). Meanwhile, the length of the two anti-parallel $\beta$-strands in caleosin (five residues in each strand; $0.35\ \text{nm} \times 5 = 1.75$ nm) is comparable to the depth (2 nm) of a monolayer of PLs. Presumably, two pairs of relatively hydrophilic residues in the anti-parallel $\beta$-strands of caleosin are located in the hydrophilic head region while the remaining three pairs of residues are embedded in the hydrophobic tail portion of the PL layer.

In the C-terminal hydrophilic domain, a putative tyrosine kinase phosphorylation site is conserved in all the five homologous sequences and located near the junction with the central hydrophobic domain (Fig. 1). In addition, five potential casein kinase II phosphorylation sites are predicted in the C-terminal domain of sesame caleosin; three of them are conserved in the five known sequences. An invariable cysteine residue prior to a casein kinase II phosphorylation site occurs near the end of the protein. Since $\beta$-mecaptoethanol treatment of total oil body proteins did not cause any changes in migration of caleosin polypeptide in SDS-PAGE, there appears to be no intra- or inter-disulfide linkage in caleosin on the surface of oil bodies (data not shown).

## Discussion

A cDNA sequence encoding caleosin, a novel oil-body protein other than the well-studied oleosin, was obtained and confirmed via immunological recognition of its over-expressed protein. Indeed, homologous sequences of one rice cDNA, one soybean cDNA, and two Arabidopsis genomic clones have been reported earlier (Accession no. X89891, AF004809, AC002332, and AF067857, respectively). Except the rice cDNA clone, no experimental result other than sequence has been documented for the other three clones. The rice cDNA clone was obtained from germinating seeds in response to abscisic acid treatment (Frandsen et al. 1996). This gene was found expressed in embryo during seed maturation, but not in other vegetative tissues unless treated with osmotic stress, and its fusion protein over-expressed in *E. coli* was shown to bind $Ca^{2+}$ in blot binding assays. The result of binding assays is in accord with the $Ca^{2+}$ binding of caleosin either extracted from oil bodies or over-expressed in *E. coli* (Fig. 4). Meanwhile, the soybean sequence was obtained from a cDNA library constructed by mRNA of maturing seeds. The results are in agreement with the Northern hybridization of caleosin mRNA in maturing sesame seeds (Fig. 6), and indicate that caleosin is a unique oil-body protein specifically expressed in developing seeds of diverse species.

Regardless the lack of sequence homology, both caleosin and oleosin comprise three structural domains, particularly a central hydrophobic anchoring domain responsible for the association of these proteins to seed oil bodies. Though the central hydrophobic segment of caleosin is much shorter than that of oleosin, a conserved proline knot-like motif occurs in caleosin comparable to the proline knot motif in oleosin (Fig. 7C). This unique similarity between these two oil-body proteins may imply a significant role associated with the proline knot motif, such as protein folding, assembly, or targeting to oil bodies as suggested by site-directed mutagenesis (Abell et al. 1997). It will be interesting to see if other oil-body proteins, e.g., Sop2 and Sop3 in sesame, also possess a hydrophobic anchoring domain with a conserved proline knot-like motif.

To prevent organelle coalescence by steric hindrance,

the amphipathic N-terminal and C-terminal domains of oleosin are presumably associated with PLs and flatted on the surface of an oil body where a hydrophobic TAG matrix is wrapped in hydrophilic surroundings (Tzen et al. 1992). Except for the amphipathic property, the length and sequence of these two domains in diverse oleosins are not conserved. By contrast, the N-terminal calcium-binding domain and the C-terminal phosphorylation domain of caleosin are hydrophilic and well-conserved in the five available sequences. It remains to be seen whether calcium-binding of the N-terminal domain and phosphorylation of the C-terminal domain provoke caleosin specific biological function(s) other than the structural role contributed by oleosin. The putative biological function(s) may be related to mobilization of oil bodies during seed germination.

The authors wish to thank Professor Chih-Ning Sun and Ms. Miki Wang for critical reading of the manuscript, and Dr. Tien-Joung Yiu of the Crop Improvement Department, Tainan District Agricultural Improvement Station for supplying mature and fresh maturing sesame seeds. The work was supported by a grant from the National Science Council, Taiwan, ROC (NSC 89-2313-B-005-002 to JTC Tzen).

## References

Abell, B.M., Holbrook, L.A., Abenes, M., Murphy, D.J., Hills, M.J. and Moloney, M.M. (1997) Role of the proline knot motif in oleosin endoplasmic reticulum topology and oil body targeting. *Plant Cell* 9: 1481–1493.

Altschul, S.F., Warren, G., Webb, M., Eugene, W.M. and David, J.L. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

Chen, E.C.F., Tai, S.S.K., Peng, C.C. and Tzen, J.T.C. (1998) Identification of three novel unique proteins in seed oil bodies of sesame. *Plant Cell Physiol.* 39: 935–941.

Chen, J.C.F., Lin, R.H., Huang, H.C. and Tzen, J.T.C. (1997) Cloning, expression and isoform classification of a minor oleosin in sesame oil bodies. *J. Biochem.* 122: 819–824.

Chuang, R.L.C., Chen, J.C.F., Chu, J. and Tzen, J.T.C. (1996) Characterization of seed oil bodies and their surface oleosin isoforms from rice embryos. *J. Biochem.* 120: 74–81.

Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1982) The helical hydrophobic moment: a measure of the amphilicity of a helix. *Nature* 299: 371–374.

Frandsen, G., Muller-Uri, F., Nielsen, M., Mundy, J. and Shriver, K. (1996) Novel plant $Ca^{2+}$-binding protein expressed in response to abscisic acid and osmotic stress. *J. Biol. Chem.* 271: 343–348.

Huang, A.H.C. (1996) Oleosin and oil bodies in seeds and other organs. *Plant Physiol.* 110: 1055–1061.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105–132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.

Murphy, D.J. (1993) Structure, function and biogenesis of storage lipid bodies and oleosins in plants. *Prog. Lipid Res.* 32: 247–280.

Napier, J.A., Stobart, A.K. and Shewry, P.R. (1996) The structure and biogenesis of plant oil bodies: the role of the ER membrane and the oleosin class of proteins. *Plant Mol. Biol.* 31: 945–956.

Qu, R., Wang, S.M., Lin, Y.H., Vance, V.B. and Huang, A.H.C. (1986) Characteristics and biosynthesis of membrane proteins of lipid bodies in the scutella of maize (*Zea mays* L.). *Biochem. J.* 234: 57–65.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *In* Molecular Cloning: a Laboratory Manual, 2nd ed. Cold Springs Harbor Laboratory Press, New York.

Shiffer, M. and Edmundson, A.B. (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* 7: 121–135.

Slack, C.R., Bertaud, W.S., Shaw, B.P., Holland, R., Browse, J. and Wright, H. (1980) Some studies on the composition and surface properties of oil bodies from the seed cotyledons of safflower and linseed. *Biochem. J.* 190: 551–561.

Tzen, J.T.C., Cao, Y.Z., Laurent, P., Ratnayake, C. and Huang, A.H.C. (1993) Lipids, proteins, and structure of seed oil bodies from diverse species. *Plant Physiol.* 101: 267–276.

Tzen, J.T.C. and Huang, A.H.C. (1992) Surface structure and properties of plant seed oil bodies. *J. Cell Biol.* 117: 327–335.

Tzen, J.T.C., Lie, G.C. and Huang, A.H.C. (1992) Characterization of the charged components and their topology on the surface of plant seed oil bodies. *J. Biol. Chem.* 267: 15626–15634.

Tzen, J.T.C., Peng, C.C., Cheng, D.J., Chen, E.C.F. and Chiu, J.M.H. (1997) A new method for seed oil body purification and examination of oil body integrity following germination. *J. Biochem.* 121: 762–768.

Wilkins, T.A. and Smart, L.B. (1996) Isolation of RNA from plant tissue. *In* A Laboratory Guide to RNA. Edited by Krieg, P.A. 2: 21–41. Wiley-Liss Inc., New York.

Yatsu, L.Y. and Jacks, T.J. (1972) Spherosome membranes. Half unit-membranes. *Plant Physiol.* 49: 937–943.

(Received June 4, 1999; Accepted August 12, 1999)