Secondary Structures of a New Class of Lipid Body Proteins from Oilseeds*

(Received for publication, August 5, 1991)

Ming Li, Linda J. Smith‡, David C. Clark‡, Reginald Wilson‡, and Denis J. Murphy‡

From the Department of Brassica and Oilseeds Research, Cambridge Laboratory, John Innes Center for Plant Science Research, Norwich NR4 7UH, and the Agricultural and Food Research Council Institute of Food Research, Norwich Laboratory, Norwich Science Park, Colney NR4 7UA, United Kingdom

The three main isoforms of the 19-kDa lipid body proteins (oleosin) have been purified to homogeneity from embryos of rapeseed. The secondary structures of these proteins have been derived from circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy and compared with the secondary structures predicted from the primary sequences. The salient feature of the primary sequence of all oleosins is its division into three defined structural domains: a central hydrophobic domain flanked by either side by relatively hydrophilic domains, respectively. Using a variety of predictive methods based on primary amino acid sequence data, the oleosins exhibited a high probability of β-strand structure in the 70-residue central hydrophobic domain, with relatively little α-helical content. Secondary structure data derived from CD and FTIR were consistent with the predictions from primary sequence, showing that the oleosins contained about 45% β-strand and 13% α-helical structure. Under high salt conditions, a 40-kDa polypeptide was obtained from purified preparations of the 19-kDa oleosins. The 40-kDa polypeptide has a very similar secondary structure, as analyzed by CD and FTIR, to that of the 19-kDa oleosins. This polypeptide is therefore probably a dimer of the 19-kDa oleosins that is formed in high salt environments. A model of the general structure of oleosin is proposed whereby the central hydrophobic domain of the protein with a predominantly β-strand structure is embedded into the nonaqueous phase of lipid-bodies. This hydrophobic region is flanked by putative α-helical structures in the polar N- and C-terminal domains which are probably oriented at the lipid-water interface.

Lipid-storing tissues in plants contain many small lipid droplets, termed lipid bodies or oil bodies, which are surrounded and stabilized by an annulus of phospholipid and protein (1). The protein components of plant oil bodies are called oleosins and it has been shown that they are not associated with any other subcellular structures. Oleosins have recently been isolated by SDS-PAGE from several species of oilseeds (2, 3) from widely separated families. The oleosins are therefore a group of proteins involved in lipid storage in plants, and predictions from their primary amino acid sequences suggest that they possess some of the structural features of the transport apolipoproteins, which stabilize the morphologically similar lipid-protein droplets (lipoproteins) found in the circulatory systems of animals (4). It is only recently that it has been recognized that oleosins constitute a single family of structural-related interfacially active oil body-associated proteins in plants. This follows the derivation of oleosin DNA and amino acid sequences from maize (5, 6), carrot (7), radish (4), rapeseed (4), soybean, and sunflower. These proteins all act as emulsifying agents at an oil-water interface. Proteinaceous emulsifying agents are of great interest to the food industry and pharmaceutical industry, as are the mechanisms responsible for the process of emulsification both in natural and synthetic foods and the packaging of some orally administered lipophilic drug molecules.

Despite the growing importance of rapeseed as a major oilseed crop, relatively little is known about the relationship between structure and function and the possible uses of its oleosins. The oleosins of rapeseed constitute a family of at least three and possibly more proteins with apparent molecular masses in the range 19 ± 0.5 kDa. It has hitherto proved impossible to separate the oleosin isoforms either using conventional techniques of chromatography, FPLC, or SDS-PAGE. In the present study, we report the purification to homogeneity of the three major isoforms of rapeseed oleosin and the determination of their secondary structures using circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy. The structures derived from these physical studies were then compared with those predicted from analysis of the primary sequences of oleosins.

EXPERIMENTAL PROCEDURES

Materials

The following chemicals were obtained from commercial sources: Sephadex G-75 and Mono Q HR 5/5 columns from Pharmacia, Lysobacter enzymogenes endopeptidase Lys-C from Boehringer Mannheim, United Kingdom, Ampholine pH 5.0–7.0 and 3.5–10 from Sigma, Carrier Amphotrope buffer pH 5–8 from Pharmacia. All other chemicals were of analytical grade.

Purification of Oil Body Membrane Proteins—Rapeseeds were gently homogenized using a pestle and mortar. The homogenization medium contained 0.4 M sucrose, 100 mM HEPES-NaOH, 10 mM KCl, 1 mM MgCl₂, 1 mM EDTA, pH 7.5. The homogenate was then filtered through four layers of cheesecloth and centrifuged at 5000 × g for 15 min. Then the oil bodies were isolated as described by Murphy and Cummins (1). In order to delipidate the oil bodies, the purified oil bodies were extracted twice in 5 volumes of Et₂O and

---

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: (0603-52571-2544; Fax: 0603-562257.

† The abbreviations used are: SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; CD, circular dichroism; FTIR, Fourier transform infrared; FPLC, fast protein liquid chromatography; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; CHAPS, 3-(3-cholamidopropyl)dimethylammonium-1-propanesulfonic acid.

Downloaded from http://www.jbc.org/ by guest on March 24, 2020
washed three times in 5 volumes of buffer (Tris-HCl, 20 mM, pH 9.0). The resulting membrane proteins were recovered by centrifugation at 100,000 × g for 1 h. All above procedures were performed at 4 °C.

Purification of 19- and 40-kDa Oleosins

Oil body membrane oleosins solubilised by 1.5% SDS were applied to a column (90 × 2.2 cm) of Sephadex G-75 equilibrated with 30 mM Tris-HCl buffer, containing 0.2% SDS, pH 9.5, and eluted with the same buffer. Fractions of 2 ml were collected at a flow rate of 20 ml/h. Fractions containing partly purified 19-kDa oleosins were collected and dialysed against 20 mM Tris-HCl, pH 9.5, for 48 h. After extensive dialysis, the proteins were further separated by PPLC on a Mono-Q column equilibrated with 30 mM Tris-HCl, pH 9.5. Fractions of 0.5 ml were collected at a flow rate of 0.4 ml/min. After extensive washing with the equilibration buffer until the protein level was nearly zero, the proteins absorbed onto the column were eluted successively with a linear gradient of 0–1 M NaCl in the equilibration buffer. Protein profiles were monitored at 290 nm. The purity of oleosins were checked by SDS-PAGE (12%) according to Schägger and Von Jagow (8). The proteins were visualized with Coomassie Blue R-250 staining. The purified 19-kDa upper, middle, and lower fractions, as identified by their mobility in SDS-PAGE, were used in the following experiments.

Protein Composition and Concentration Analysis

Amino acid composition of protein samples for CD analysis was determined by the method of Lindroth (9). The protein samples were hydrolysed in 6 N HCl at 110 °C for 24 h, evaporated, and then free amino acids were determined by reverse phase high performance liquid chromatography of their o-phthalaldehyde-derivatives. Individual amino acids were detected by fluorescence (350-nm excitation, 450-nm emission), using a Gilson 121 fluorometer, and quantified by integration of the peaks with reference to standard curves prepared for each amino acid.

Protein concentrations were determined, in the presence of SDS, by a modified Lowry method (10) and bovine serum albumin was employed as reference protein for standard curves. The protein concentrations of the samples used in the analysis of circular dichroism were calculated from the analysis of amino acid composition as described above.

Secondary Structure Prediction from Amino Acid Sequence

The secondary structure of rapeseed 19-kDa protein was predicted from primary sequence data using several different methods based on a number of well known tertiary structures of proteins (11, 12). The hydrophobicity plot of the protein amino acid sequence was predicted by the methods of Chou and Fasman (11) and Kyte and Doolittle (12). The secondary structure of the proteins was empirically predicted based on the Chou and Fasman rules (11). A model of the conformation of oleosins on the surface of an oil body is proposed (Fig. 5) based on this prediction and on data derived from CD, FTIR, and isoelectric point analysis of the isomers of the 19-kDa proteins.

Circular Dichroism Spectroscopy

CD spectroscopy spectra were recorded in the wavelength range 260–190 nm on a Jasco J-600 A Spectropolarimeter under constant nitrogen flush at 21 °C. Data were recorded on-line using an IBM personal computer. Spectra presented are the average of four scans, recorded at 10 nm/min, using silica quartz spectrophotometer cells generally of 0.1- and 1-mm path length. An instrument sensitivity of ±20 mdeg full scale was routinely used, along with a 4-s time constant (13). The instrument calibration was regularly checked during the course of the work using ammonium d-10-camphorsulphonate (14) and d-(-)-pantolactone (15).

Calculation of Percentage Secondary Structures from CD Data

The CD data were expressed as mean residue ellipticity ( Φ ) (deg·cm²·mol⁻¹). The mean residue weights of samples were determined from analysis of amino acid composition. Data at appropriate intervals between 260 and 190 nm were analyzed by standard linear combination methods. The principle of this technique involves fitting the experimental spectrum with basis spectra for secondary structure motifs by a least-square curve fitting method. The basis spectra are prepared from reference spectra obtained from soluble proteins of known three-dimensional structure determined by X-ray crystallography. Two widely accepted methods were used. The first was the CONTIN program (16) which contains an internal data base containing spectra of 15 reference proteins. This program was used in unmodified form and analysis of the experimental spectra was conducted between 240 and 190 nm at 1-nm intervals. The program returns estimates for the fractions of α-helix, total β-structure (i.e., sheet and turn), and unordered structure present in the sample. A sum of structures constraint is also used such that the sum of the fractions for all the secondary structures should be 1.0. The second method was due to Chen (17, 18), but the data base of reference spectra used was enlarged to 15 using published data (19). In addition, this program returned estimates for the fractions of α-helix, β-sheet, β-turn, and unordered secondary structure motifs.

Fourier Transform Infrared Spectroscopic Studies

FTIR spectra of 19-kDa proteins in solid state were obtained at room temperature using a Bio-Rad FT580 spectrometer using a liquid nitrogen cooled mercury-cadmium telluride (MCT) detector at a resolution of 2 cm⁻¹. Triangular apodization was employed. The samples were presented to the spectrometer as potassium bromide pellets. A small amount (circa. 0.2 mg) was mixed with finely ground potassium bromide (IR grade) and pressed for 5 min at 10 tonnes. A total of 1200 scans were collected for each spectrum. A potassium bromide disc was used for the estimation of background.

The conformational assignments of amide I, II, and III bands in FTIR spectra followed in this study were adopted from Renugopakrishnan et al. (20).

(a) α-Helical Structure—The amide I and III frequencies of α-helical structures occur in the 1645–1659- and 1622–1630-cm⁻¹ region, respectively (20).

(b) β-Strand Structures—The assignments proposed by Chirgadze and Nevskaya (23, 24) were followed. Antiparallel β-strand structures manifest A and B, components in the 1670–1690-cm⁻¹ region in IR spectra and a B-component around 1629 cm⁻¹ (23). The parallel-strand structures manifest A, B-components around 1640 cm⁻¹ (24).

Amide III frequencies occur in the 1290–1245-cm⁻¹ region (21).

(c) β-Turn Structures—The amide I vibrations of β-turn structures have been the subject of detailed studies in the last decade. The assignments were adopted from Lagant et al. (25–27), Sexton (28), and Ishizaki et al. (29). The type I β-turn structures occur around 1658, 1663, and 1236 cm⁻¹. Type II β-turn structures occur around 1654, 1662, 1684, and 1255 cm⁻¹. Type III β-turn structures occur around 1658, 1650, and 1277 cm⁻¹.

(d) Unordered Structures—Although it is difficult to assign a particular subregion of amide I region to unordered structures, the 1654–1657-cm⁻¹ region appears to characterize unordered structure on the basis of IR study of feather keratin and denatured proteins (30).

(e) Mixed Domains of Extensive β-Turns and Less Extensive β-Strand Structures—From well established β-strand vibrations and newly established β-turn vibrations, the characteristic vibrations for a protein containing mixed motifs should be expected to occur in 1670–1680 and around 1240 cm⁻¹ (31).

In preparation for curve fitting, band center frequencies were conformed by second derivative spectroscopy. Secondary structure components were accomplished by least-squares iteration. Gaussian band shapes were assumed for the deconvoluted components.

RESULTS

Purification of Oleosins—In rapeseed, the 19-kDa family of oleosin proteins can constitute up to 80% of the total protein associated with theflowering oil body fraction, obtained from low speed centrifugation of a seed homogenate (32). The majority of the lipid in this fraction was readily removed by ether extraction, and the “delipidated” oil body membrane fraction, highly enriched in oleosins, was the starting material for subsequent purification. Neutral lipid-associated proteins are commonly subjected to mild delipidation with diethyl ether or 1.5% SDS and dialysed against 20 mM Tris-HCl, 20 mM, pH 9.0, for 1 h. All above procedures were performed at 4 °C. The resulting membrane proteins were recovered by centrifugation at 100,000 × g for 1 h. All above procedures were performed at 4 °C.
G-75. Two major protein peaks were obtained. Both peaks were somewhat enriched in 19-kDa oleosins but the first peak also contained numerous other membrane proteins, whereas the second peak contained only small amount of non-oleosin bands. Therefore, the fractions from the second peak were collected and dialyzed against 20 mm Tris-HCl, pH 9.5, for 48 h. After extensive dialysis, the proteins were further separated by FPLC on a Mono Q column (Fig. 1). Several peaks were eluted from this column, and their purity was checked by performing SDS-PAGE on all fractions. The 19-kDa oleosins were resolved into three bands corresponding to peaks I, II, and III of the FPLC profile shown in Fig. 1A. The other peaks shown in the FPLC chromatogram contained non-19-kDa oleosin proteins or mixtures of the oleosins. Analysis of peaks I, II, and III by SDS-PAGE showed that each peak contained a single homogeneous 19-kDa oleosin isoform as shown in Fig. 1B. The oleosin isoforms were of similar but slightly different molecular masses and for convenience are termed 19-kDa upper band (peak I), middle band (peak II), and lower band (peak III). In addition to the three 19-kDa oleosin bands resolved by FPLC, a further fraction (Fig. 1A, peak IV) was obtained which appeared almost exclusively as a 40-kDa band on an SDS-PAGE gel (Fig. 1C). This band was not present in the original eluate from the G-75 column.

**Fig. 1.** A, Mono Q column chromatography. Fractions of 0.5 ml were collected at a flow rate of 0.4 ml/min. Peaks I, II, III, and IV indicate the fractions which were taken as the final preparations. B and C, SDS-PAGE (13.5%) for checking the purity of fractions from Mono Q column chromatography. Lane 1 of B, molecular mass markers; lanes 2–4 of B, 19-kDa upper, middle, and lower bands which represent peaks II, III, and I from Mono Q column chromatogram, respectively. Lane 1 of C, molecular mass markers; lanes 2–4, 40-kDa oleosin bands from fractions in peak IV of Mono Q column chromatogram.

**Determination of Secondary Structures of Oil Body Proteins**

SDS-PAGE (13.5%) for checking the purity of fractions from Mono Q column chromatography. Fractions of 0.5 ml were collected at a flow rate of 0.4 ml/min. Peaks I, II, III, and IV indicate the fractions which were taken as the final preparations. B and C, SDS-PAGE (13.5%) for checking the purity of fractions from Mono Q column chromatography. Lane 1 of B, molecular mass markers; lanes 2–4 of B, 19-kDa upper, middle, and lower bands which represent peaks II, III, and I from Mono Q column chromatogram, respectively. Lane 1 of C, molecular mass markers; lanes 2–4, 40-kDa oleosin bands from fractions in peak IV of Mono Q column chromatogram.

In addition to the three 19-kDa oleosin bands resolved by FPLC, a further fraction (Fig. 1A, peak IV) was obtained which appeared almost exclusively as a 40-kDa band on an SDS-PAGE gel (Fig. 1C). This band was not present in the original eluate from the G-75 column.

**Fig. 1.** A, Mono Q column chromatography. Fractions of 0.5 ml were collected at a flow rate of 0.4 ml/min. Peaks I, II, III, and IV indicate the fractions which were taken as the final preparations. B and C, SDS-PAGE (13.5%) for checking the purity of fractions from Mono Q column chromatography. Lane 1 of B, molecular mass markers; lanes 2–4 of B, 19-kDa upper, middle, and lower bands which represent peaks II, III, and I from Mono Q column chromatogram, respectively. Lane 1 of C, molecular mass markers; lanes 2–4, 40-kDa oleosin bands from fractions in peak IV of Mono Q column chromatogram.

**Estimation of the Secondary Structure Content by CD**—The CD spectra of the oleosin fractions are plotted together in Fig. 3 for ease of comparison. The general features of the spectra were similar but there were some subtle differences. The spectrum from the upper fraction was the most different in terms of band structure and intensity. The other spectra showed consistent band structure with the exception of the 40-kDa fraction at wavelengths below 205 nm. The general features of the spectra were consistent with those of proteins containing significant levels of β-structure since the negative band extended down to a trough at 206 nm. In addition, the intensity of the negative CD band was too weak for the protein to contain more than 20% α-helix. However, the spectra were sufficiently structured to indicate that delipidation of the protein during preparation does not destroy the secondary structure of the protein. It is possible that tightly bound lipid remaining after delipidation is an important factor in maintaining the oleosin secondary structure.

A summary of the results of the analysis of the CD spectra for secondary structure content is shown in Table I. The results of both analysis methods were in approximate agreement, showing that the samples investigated had a similar secondary structure content of approximately 20–15% α-helix, 60–65% β-structure, and 20–30% unordered structure. The β-structure content was further subdivided into approximately 40–48% β-strand and 12–17% turn by the Chen method (18).
FIG. 2. Secondary structure prediction from the full-length primary amino acid sequence of the rapeseed oleosin Bn-III. A, secondary structure prediction and hydropathy plot of the amino acid sequence of the rapeseed oleosin Bn-III, using the methods of Chou and Fasman (11) and Kyte and Doolittle (12). The major feature is the division of the sequence into three defined domains, i.e. a central hydrophobic domain from residue 60-131 flanked on either side by relatively hydrophilic domains from residues 1-59 and 132-195, respectively. The central hydrophobic domain corresponds with the region of predicted $\beta$-strand structure. B, hydrophobicity and hydrophilicity plot for 19-kDa oleosin as a function of the residue number (Chou-Fasman). Notice that a hydrophobic domain in the middle is flanked on either side by relatively hydrophilic domains. C, surface probability predicted by using Chou-Fasman prediction. Notice that the main surface
probability is on the either side of hydrophilic domains, especially in the N-terminal domain. D, antigen index predicted by using Chou-Fasman prediction. The prediction here is that the antigenic sites are all on the hydrophilic domains. In B-D the predicted β-strand structure (\(\sim\)) lies mainly in the central, hydrophobic domain.
Examination of each method separately indicated that structure determinations using the CONTIN program (16) showed more variation between the samples. The upper isoform was characterized by approximately 5% less α-helix than the other two isoforms. The intensity differences between the middle and lower isoform spectra were rationalized by alteration in the relative amounts of β-sheet and unordered structure. The 40-kDa fraction was distinguishable from the other samples by a 7% reduction in α-helix content. In contrast, the Chen analysis returned similar structure contents for middle, lower, and 40-kDa samples. The upper isoform was characterized by an altered β-sheet to β-turn ratio.

Conformational Assignments of Proteins from FTIR Spectra—Typical FTIR spectra of the oleosins in the solid state are shown in Fig. 4. All panels of Fig. 4, ranging from 1000 to 2000 cm⁻¹, show the deconvolution FTIR spectra of the amide I, I1, and I11 regions, respectively. The frequencies of the main peaks and their tentative assignments are listed in Table II.

The main amide I peaks of the 19-kDa upper band are shown in Fig. 4A as a broad doublet with two main peaks at 1662.5 and 1654.5 cm⁻¹, together with two small peaks at 1633, 1610 and shoulders at 1684.5 and 1670.5 cm⁻¹. The main amide II peaks are found at 1545, 1535, 1516 with a shoulder at 1526 cm⁻¹. The amide III peaks occur at 1263, 1252, 1243, and 1232 cm⁻¹. For the 19-kDa middle band as shown in Fig. 4B, the main amide I peak occurs at 1658 with shoulders at 1687 and 1679 cm⁻¹ and two smaller peaks at 1631 and 1609.5 cm⁻¹. The amide II peaks occur at 1542, 1524.7, and 1517 cm⁻¹ with a shoulder at 1534 cm⁻¹. The amide III peaks occur at the same frequencies as in the 19-kDa upper band. For the 19-kDa lower band, as shown in Fig. 4C, amide I peaks occur at 1662, 1634, 1609.5 with shoulders at 1672 and 1658 cm⁻¹. The amide II peaks occur at 1550, 1542, and 1515 cm⁻¹. The amide III peaks again occur at the same frequencies as the 19-kDa upper band. For the 40-kDa bands, as shown in Fig. 4D, the main amide I peaks are shown as a broad triplet with peaks at 1673, 1663, and 1655, together with a small peak at 1634 and shoulder at 1685 cm⁻¹. The main amide II peaks are found at 1634, 1550, 1541, and 1518 cm⁻¹. The amide III peaks occur at 1261, 1250, and 1231 cm⁻¹. Although the FTIR spectrum of the 40-kDa protein is slightly different it is still quite similar to the spectra of 19-kDa proteins with regard to the peak positions and peak areas.

Based on numerous observations on synthetic polypeptides and proteins (20–31), the broad peaks in FTIR deconvoluted spectra of the three oleosin isoforms at 1663 (upper band), 1659 (middle band) and 1662 cm⁻¹ (lower band), respectively, are indicative of β-sheet structure based on amide I peaks of polyserine, 1668 cm⁻¹ (36); carbonic anhydrase (40% β-strand), 1668 cm⁻¹; bovine pancreatic trypsin inhibitor (37% β-strand), 1667 cm⁻¹ (28); and our CD data. The shoulders at 1684.5 and 1687 cm⁻¹, respectively, are attributable to anti-parallel β-strand structures (25). Shoulders at 1670.5, 1679, and 1672 cm⁻¹ are indicative of parallel β-strand structure. The bands at 1633, 1631, and 1634, respectively, are indicative of β-strand structure B, component (23, 24). Bands at 1654.5, 1659, and the shoulder at 1658 cm⁻¹ could be α-helical or unordered structures (21).

In general, the amide II frequencies for different secondary structures are less well established than amide I and III frequencies. The amide II bands show two components at 1535, 1542, and 1550 cm⁻¹ characteristic of α-helical and β-strand structures. The amide III bands occur at the same frequencies at 1263, 1252, and 1232 cm⁻¹ for all three samples. Bands at 1263 and 1252 cm⁻¹ are indicative of β-strand or mixed domains of extensive β-turns and less extensive β-strand structures. It is quite difficult to assign a particular subregion of amide I, II, and III regions to unordered structures.

The percentages of secondary structures of all samples were calculated from FTIR data. Corresponding values for the secondary structure of these proteins were provided for comparison with the results derived from CD and primary structure prediction (Table I).

**DISCUSSION**

In the present study, three isoforms of the lipid-associated 19-kDa oleosins from rapeseed were purified to homogeneity.
Determination of Secondary Structures of Oil Body Proteins

In addition, a 40-kDa protein, which is probably a dimer of 19-kDa oleosins, was isolated. In order to effect this purification, it was found that it was necessary to add at least 1.5% (w/v) SDS to the crude oleosin preparation in order to obtain complete solubilization. Non-ionic detergents, such as CHAPS and other ionic detergents, such as cholic acid or deoxycholic acid, were not suitable solubilising agents. Following gel-filtration chromatography in the presence of 0.2% SDS, the desired fractions were collected and dialyzed extensively against Tris-HCl buffer in order to obtain a differential charge distribution on the surface of the various proteins in each fraction and allow further purification by ion-exchange column chromatography. Using a Mono Q column, well separated fractions containing the three homogeneously purified oleosin isoforms were obtained. In the absence of such extensive dialysis, only a single protein peak was obtained from the Mono Q column.

Oleosins are hydrophobic proteins which are aligned at a lipid-water interface. They are therefore quite different to soluble proteins and to integral membrane proteins as regards the constraints that will determine their secondary structure. Soluble proteins are surrounded by a polar aqueous milieu, whereas integral membrane proteins normally span an amphipathic lipid bilayer and hence have two opposite faces exposed to different aqueous compartments and a hydrophobic core which is embedded in the membrane interior. The algorithms used for the prediction of protein secondary structures from primary sequences are based largely upon data from soluble globular proteins. One must therefore exercise caution in the use of such predictions with unusual hydrophobic proteins, such as the oleosins. For this reason, it is important to employ physical techniques, such as CD and FTIR spectroscopy, in addition to predictive methods in order to elucidate more fully the structure of the oleosins. However, it is important to appreciate the strengths and limitations of

FIG. 4. Infrared Fourier self-deconvolution spectra of the upper (A), middle (B), and lower (C) 19-kDa oleosins and 40-kDa (D) polypeptide in solid states.

| Table II | Fourier transform infrared frequencies of oleosins in solid state |
|----------|--------------------------------------------------|
|          | Upper | Middle | Lower | 40 kDa | Assignments            |
| 1635(sh)*| 1637(sh) | 1636(sh) | 1638(sh) | 1635(sh) | Antiparallel β-strand |
| 1671(sh) | 1677(sh) | 1672(sh) | 1673    | β-Strand or mixed domains |
| 1662     | 1658    | 1662    | 1663    | Type II β-turn |
| 1655     | 1650    | 1658(sh) | 1655    | Type III β-turn or α-helix |
| 1633     | 1631    | 1634    | 1634    | β-Strand |
| 1546     | 1543    | 1550    | 1550    | α-Helix |
| 1516     | 1517    | 1515    | 1518    | β-Strand |
| 1263     | 1263    | 1263    | 1261    | β-Turn repetitive |
| 1252     | 1252    | 1252    | 1250    | β-Turn |
| 1243     | 1243    | 1243    | 1243    | β-Turn |
| 1232     | 1232    | 1232    | 1231    | β-Strand |

*All (sh) represent shoulders in FTIR spectra.
Determination of Secondary Structures of Oil Body Proteins

![Diagram of oil body protein structure](http://example.com/diagram.png)

**Fig. 5. Proposed model of the orientation of the rapeseed oleosin, Bn-III, at lipid-water interface of a storage oil body.**

A. N-terminal amphipathic \( \alpha \)-helical domain with some \( \beta \)-turn and random structures; B, central hydrophobic domain possibly made up of \( \beta \)-strand structure embedded in the nonaqueous core of the oil body and possibly similar to putative \( \beta \)-strand repeats of apolipoprotein B-100 (40, 43); C, C-terminal polar domain possibly containing a small amount of amphipathic \( \alpha \)-helical structure with some \( \beta \)-turn and random structures.

It was reported that CD data measured to 190 nm have an information content of nearly 4 (37). This is equivalent to the number of variables present in the CONTIN program. It is not always necessary for the information content to be equal to or greater than the number of variables, especially if one of the reference spectra used is identical to the measured spectrum. However, this is not the case here. The structural feature determined with the greatest degree of reliability from CD is the \( \alpha \)-helix content. Analysis by the CONTIN method identified that both upper and 40-kDa fraction differed in \( \alpha \)-helix content compared with the middle and lower fractions. Such differences were not as pronounced when the data were analyzed using the Chen method. However, since this method used five variables it may be criticized as the interpretation of the data. It is necessary to continue data collection down to 184 nm to increase the information content of the spectra to 5 (38). Nevertheless, these results are worthy of note because they show excellent agreement with the FTIR data. The accuracy of estimation of \( \beta \)-structure by CD is rather poor especially if spectra are confined to >190 nm (38). Given this limitation there is reasonable agreement between the total \( \beta \)-structure returned by CONTIN and the combined \( \beta \)-sheet and turn returned by the Chen analysis.

The results of the secondary structure analysis by CD and FTIR have identified small but measurable differences between the oleosin fractions studied here. The origin of these differences is unclear. One possible explanation could involve differences in tightly bound lipid. Alternatively, perhaps the structure of the 19-kDa upper band is more constrained than the other two isoforms. The differences detected in the structure of the 40-kDa fraction may reflect structural changes that occur upon dimerization.

The predicted secondary structure of oleosins is of great interest in that it includes a 71-residue continuous hydrophobic domain which has a very high probability of \( \beta \)-strand structure according to all the predictive methods that were used both here and elsewhere (4). Hydrophobic \( \beta \)-strand structures are most unusual, although smaller \( \beta \)-strand domains are believed to exist in the membrane-immersed domains of bacterial porins (39) and, possibly, in the lipid-binding domains of apolipoprotein B-100 in low density lipoproteins (40, 41). Some sequence similarity between oleosins and animal apolipoproteins has already been noted, particularly with regard to a putative amphipathic \( \alpha \)-helical domain which would stabilize a phospholipid-water interface (4). Like oleosins, apolipoproteins exist at a single lipid-water interface and are responsible for the stabilisation of neutral lipid-phospholipid droplets. The analogy between the two types of protein must not, however, be taken too far. Although apolipoproteins stabilize relatively short-lived (\( t_{1/2} = 0.5-2 \) h) animal serum transport lipoproteins, oleosins stabilize very long-lived (\( t_{1/2} = \) many years) storage oil bodies in the dry seeds of plants.

The secondary structures of the oleosins which were derived from CD spectra correspond closely with the predictions from the full-length primary sequence of the oleosin isoform Bn-III as shown in Fig. 2A and previously published as an incomplete sequence termed nap-II (4). According to these predictions, the oleosins contain a high \( \beta \)-strand content (about 44%) with relatively little \( \alpha \)-helical structure (12%). Combining the CD data and the primary sequence predictions, one can derive an oleosin structure whereby the central hydrophobic domain is dominated by \( \beta \)-strands, with the flanking N- and C-terminal polar domains containing a combination of random structure, turns, and a relatively small amount of \( \alpha \)-helix. The three isoforms of the 19-kDa oleosins exhibited very similar CD spectra, and their derived secondary structures are correspondingly similar. Hence, \( \beta \)-strand structure accounts for 40, 46, and 47%, respectively, of the upper, middle, and lower 19-kDa bands, whereas the \( \alpha \)-helical contents were 12, 13, and 12%, respectively. From the primary sequence, the length of the central hydrophobic domain is about 36% of the full-length of the protein, which is very close to the proportion of \( \beta \)-strand structure derived from both CD and FTIR measurements and from secondary structure predictions. This central hydrophobic domain is highly conserved in all 12 oleosin sequences derived to date from oilseeds as diverse as carrot (7), radish (4), rapeseed (4), maize (5, 6), soybean and sunflower and therefore probably represents an important structural motif in this class of proteins. In contrast, the polar flanking domains are relatively variable in their sequence. The C-terminal regions of all oleosins contain potential amphipathic \( \alpha \)-helical domains of about 15–20 residues but are otherwise rather dissimilar to one another. The N-terminal regions are even more variable, even among isoforms from the same plant species such as rapeseed (4). Such heterogeneity in the polar terminal regions may give rise to differences in the extent of lipid binding and the observed differences in antigenic determinants between the various oleosins (35).

Based on the above investigation and on previous studies (4), the following general model for the secondary structure of oleosins can be proposed (see Fig. 5). The oleosins are made up of a hydrophobic domain in the middle part of the protein from residues 60–131 flanked on either side by relatively hydrophilic domains from residues 1–59 and 132–195, respectively. The central hydrophobic domain corresponds with the region of predicted and measured \( \beta \)-strand structure and is probably embedded in the nonaqueous core of the oil bodies. The finding that 19-kDa oleosins can apparently form 40-kDa dimers with very little change in secondary structure raises the possibility that the H-bond potential of this large \( \beta \)-strand region may be satisfied by dimeric, or even oligomeric, associations. This would enable the protein to penetrate deeply into the lipid phase rather than being aligned at the aqueous interface. This hydrophobic region is probably flanked by some \( \alpha \)-helical, turn, and unordered structures in...
the polar N- and C-terminal domains. The α-helical structures in both the N- and C-terminal regions are potentially amphipathic and are probably oriented at the lipid-water interface. Positively charged residues in these domains could then interact with phospholipid on the surface of oil bodies in a similar manner to that proposed for many serum apolipoproteins (42, 43).

The N-terminal domain of the oleosin, Bn-III, contains a predicted 18-residue α-helical region in which the 8 residues proximal to the N terminus are polar, whereas the 10 residues distal to the N terminus are nonpolar or hydrophobic (44). Hence this domain may be oriented perpendicular to the oil-water interface with 8 residues extending into the aqueous phase, whereas 10 residues extend into the lipid phase. Interestingly, this putative amphipathic α-helix contains a similar manner to that proposed for many serum apolipoproteins (42, 43). An analogous role has been proposed for oleosins in oilseeds during post germinative lipid mobilization (3, 5).

Acknowledgments—We thank Dr. J. G. Turner for analysis of amino acid composition. We also thank Drs. M. J. Hills, S. Rawsthorne, and J. Ross for useful discussions about the manuscript.

REFERENCES
1. Murphy, D. J., and Cummins, I. (1981) Phytochemistry (Oxf.) 20, 2063–2069
2. Qu, R., Wang, S., Lin, Y., Vance, V. B., and Huang, A. H. C. (1986) Biochem. J. 235, 57–65
3. Murphy, D. J., Cummins, I., and Kang, A. S. (1989) J. Sci. Food Agric. 48, 209–223
4. Murphy, D. J., Keen, J. N., O’Sullivan, J. N., Au, D. M. Y., Edwards, E.-W., Jackson, P. J., Cummins, I., Gibbons, T., Shaw, C. H., and Ryan, A. J. (1991) Biochim. Biophys. Acta 1088, 86–94
5. Vance, V. B., and Huang, A. H. C. (1987) J. Biol. Chem. 262, 11275–11279
6. Qu, R., and Huang, A. H. (1989) J. Biol. Chem. 265, 2238–2243
7. Hatzopoulos, P., Franz, G., Choy, L., and Sung, R. Z. (1990) Plant Cell 2, 457–467
8. Schägger, H., and Von Jagow, G. (1987) Anal. Biochem. 166, 368–379
9. Lindroth, P., and Mopper, K. (1979) Anal. Chem. 51, 1667–1674
10. Markwell, M. A. K., Hass, S. K., Tolbert, N. E., and Beiber, L. L. (1988) Methods Enzymol. 72, 269
11. Chou, P. Y., and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276
12. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
13. Clark, D. C., and Smith, L. J. (1989) J. Agric. Food Chem. 37, 627
14. Takakura, T., Konno, T., and Meguro, H. (1975) Anal. Sci. 1, 215–218
15. Konno, T., Meguro, H., and Tuzimura, K. (1975) Anal. Biochem. 67, 226–232
16. Provencher, S. W., and Gluckner, J. (1981) Biochemistry 20, 33–37
17. Chen, Y.-H., Yang, J. T., and Martinez, H. M. (1972) Biochemistry 11, 4120–4131
18. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) Biochemistry 13, 3530–3539
19. Yang, J. T., Wu, C.-S. C., and Martinez, H. M. (1986) Methods Enzymol. 130, 268–269
20. Renugopalakrishnan, V., Rapaka, R. S., Collette, T. W., Carreira, L. A., and Bhatnagar, R. S. (1985) Biochem. Biophys. Res. Commun. 126, 1029–1035
21. Parker, F. S. (1983) Application of IR, Raman, and Resonance Raman Spectroscopy in Biochemistry, Plenum Press, New York
22. Krissin, S., and Dwivedi, A. M. (1982) Science 216, 407–408
23. Chirgadze, Yu. N., and Nevskaya, N. A. (1976) Biopolymers 15, 607–625
24. Chirgadze, Yu. N., and Nevskaya, N. A. (1976) Biopolymers 15, 627–636
25. Renugopalakrishnan, V., Hozowitz, P. M., and Glimcher, M. J. (1985) J. Biol. Chem. 260, 11406–11413
26. Lagart, P., Vergoten, G., Fleury, G., and Loucheux-Lefèvre, M. H. (1984a) Eur. J. Biochem. 139, 137–148
27. Lagart, P., Vergoten, G., Fleury, G., and Houde, M. H. (1984b) Eur. J. Biochem. 139, 149–154
28. Seaton, B. A. (1985) Biophysical Studies of Protein Conformation. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA
29. Ishizaki, H., Balaram, P., Nagaraj, R., Venkatachalanpathi, V. Y., and Tu, A. T. (1981) Biochim. Biophys. Acta 659, 509–517
30. Koenig, J. L., and Tabb, D. L. (1980) In Analytical Applications of FT-IR to Molecular and Biological Systems (Durig, J. R., ed) pp. 241–266, D. Reidel Publishing Co., Dordrecht, The Netherlands
31. Ishizaki, H., McKay, R. H., Norton, T. R., Yasunobu, K. T., Lee, J., and Tu, A. T. (1979) J. Biol. Chem. 254, 9661–9665
32. Murphy, D. J., Cummins, I., and Kang, A. S. (1989) Biochem. J. 265, 289–295
33. Murphy, D. J., and Mukherjee, K. D. (1987) Lipids 22, 293–298
34. Swaney, J. B. (1980) J. Biol. Chem. 255, 8791–8797
35. Au, D. M. Y., Kang, A. S., and Murphy, D. J. (1989) Arch. Biochem. Biophys. 273, 516–526
36. Koenig, J. L., and Sutton, P. L. (1971) Biopolymers 10, 89–96
37. Johnson, W. C., Jr. (1988) Annu. Rev. Biochem. Chem. 17, 145–166
38. Johnson, W. C., Jr. (1990) Proteins Struct. Funct. Genet. 7, 205–214
39. Jahnig, F. (1990) Trends Biochem. Sci. 15, 93–95
40. Knott, T. J., Pease, R. J., Powell, L. M., Wallis, S. C., Roll, S. C., Innerarity, T. L., Blackhart, B., Taylor, W. H., Marcel, Y., Milne, R., Johnson, D., Fuller, M., Lusis, A. J., McCarthy, B. J., Mahley, R. W., Levy-Wilson, B., and Scott, J. (1986) Nature 323, 734–738
41. Li, W. H., Sparrow, D. A., De Loof, H., Rosseneu, M., Lee, F. S., Goto, Z. W., and Kan, A. M., and Chan, L. (1986) Nature 323, 738–741
42. Segrest, J. P., Jaelsen, R. L., Morrisset, J. D., and Gotto, A. M. (1974) FEBS Lett. 24, 247–253
43. Segrest, J. P. (1977) Chem. Phys. Lipids 18, 7–22
44. Li, W.-H., Tanimura, M., Luo, C. C., Datta, S., and Chan, L. J. (1988) J. Lipid Res. 29, 245–271
45. Kinnamon, P. K., Jackson, P. L., Smith, L. C., Gotto, A. M., and Sparrow, J. T. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 4848–4851
Secondary structures of a new class of lipid body proteins from oilseeds.
M Li, L J Smith, D C Clark, R Wilson and D J Murphy

J. Biol. Chem. 1992, 267:8245-8253.