Expression and Characterization of the N-terminal Domain of an Oleosin Protein from Sunflower*

(Ming Li, James S. Keddie, Linda J. Smith†, David C. Clark‡, and Denis J. Murphy§)

From the Department of Brassica and Oilseeds Research, John Innes Centre for Plant Science Research, Norwich NR4 7UH, United Kingdom

Oil bodies of plant seeds contain a triacylglycerol matrix surrounded by a monolayer of phospholipids embedded with alkaline proteins termed oleosins. Although oleosins are amphipathic proteins, they are unlike bilayer membrane proteins since they are associated with a single lipid:water interface at the oil body surface. Oleosins are unusual proteins because they contain a 70–80-residue uninterrupted nonpolar domain, flanked by relative polar C- and N-terminal domains. In the present study, we report the expression of the N-terminal domain of the 18-kDa oleosin isoform from sunflower as a recombinant fusion protein in Escherichia coli and the determination of its secondary structure using CD and Fourier transform infrared spectroscopy either as a purified but partially denatured peptide or reconstituted into liposomes. The structure derived from physical studies was then compared and assigned with those predicted from analysis of the primary sequence of the N-terminal domain. Based on data derived from CD spectroscopy analysis of purified and partially renatured N-terminal polypeptide, it contains about 10% α-helical structure, 20–30% β-strand structure, approximately 8% β-turn structure, and 60% random coil structure. However, analysis of the polypeptide reconstituted into liposomes showed an increased content of α-helical structure to about 20% and an increased β-strand structure content to about 30–40%. Data derived from Fourier transform infrared spectroscopy studies and compared with the data predicted from the primary sequence showed the peptide is well structured with some antiparallel β-strand structure from residues 2–9, parallel β-strand structure from residues 30–37 and/or 42–49, and α-helical structure from residues 10–23 and/or 43–49. There is potential amphipathic α-helix from residues 10–23. Based on these results, the following model for the secondary structure of the N-terminal domain of sunflower oleosin can be proposed. Residues 2–9 would produce amphipathic antiparallel β-strand structure. Residues 10–23 would produce an amphipathic α-helical structure. Residues 30–37 and/or 42–49 would give parallel β-strand structure, or residues 42–49 could form a nonpolar α-helical structure that would insert into the oil matrix.

Oleosins are a novel class of amphipathic proteins, which are specifically associated with lipid storage bodies in the seed tissues of plants (1–5). In high oil-containing seed tissues such as rapeseed embryos, oleosins may constitute up to 20% of the total cell protein at maturity (6). The protein has been shown localized exclusively on the surface of the lipid storage bodies in seed tissues (2, 7–9). It is believed that the major function of oleosin is to stabilize the small (less than 1 μm) lipid storage bodies in seed tissues, particularly during seed dehydration and subsequent rehydration following germination. A further function of the oleosin molecule may be to serve as a lipase binding site to increase the efficiency of lipolysis during mobilization of storage lipids (1, 10).

Although oleosins are amphipathic proteins, they are unlike bilayer membrane proteins since they are associated with a single lipid:water interface at the oil body surface. Oleosins are therefore analogous to other monolayer-associated proteins, such as those involved in lipid storage and transport in the circulatory systems of animals. Indeed, it has been pointed out that oleosins share some structural features with animal apolipoproteins, including a conserved amphipathic α-helical region (1, 10, 11). Oleosins appear to act as emulsifying agents at an oil:water interface. Such proteinaceous emulsifying agents are of great interest to the food and pharmaceutical industries, as are the mechanisms responsible for the process of emulsification both in natural and synthetic foods and in the packaging of some orally administered lipophilic drug molecules.

Oleosins are unusual proteins because they contain a 70–80-residue uninterrupted nonpolar domain, flanked by relatively polar C- and N-terminal domains (1, 10–14). According to secondary structure predictions based on the primary sequences of the 21 oleosins available to date, the oleosin molecule is likely to consist of a central hydrophilic β-strand structure embedded in the nonaqueous phase of the lipid bodies (10, 11). This hydrophilic region is flanked by polar N- and C-terminal domains with some putative α-helical structures. The C-terminal α-helix is amphipathic and is probably oriented at the lipid:water interface, extending into the aqueous phase (10). The N-terminal domain has hydrophilic and hydrophilic residues distributed throughout its sequence and has been predicted to contain some secondary structure, although it has not yet been possible to make specific structural assignments with any confidence in this domain. It is therefore assumed that the amphipathic parts of the N-terminal sequence interact with the lipid:water interface of the oil body at many locations. In a recent structural study utilizing circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy, it was demonstrated that the

*Funding was provided by a grant from the Agricultural & Food Research Council, United Kingdom (to D. J. M.). 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.

† Present address: AFRC Institute of Food Research, Norwich Laboratory, Colney, Norwich NR4 7UH, U.K.

‡ To whom correspondence should be addressed. Tel.: 44-603-52571; Fax: 44-603-259882.

1The abbreviations used are: FTIR, Fourier transform infrared; PBS, phosphate-buffered saline; PC, 1-stearoyl-2-oleoyl-sn-glycerol-3-phosphocholine; TX-100, Tween-80; TFE, trifluoroethanol; TFE, trifluoroethanolamine.
three major oleosin isoforms of rapeseed contained, as predicted, about 45% β-strand structure and about 12% α-helical structure (11). Although the percentages of secondary structure of full-length oleosin is now known (11), the assignment and distribution of these secondary structures in each of the three domains have yet to be determined. Many of the α-helical and β-strand structures were assigned to the central and C-terminal domains in the previous study. This left open the question of whether the 50-residue N-terminal domain was largely made up of random coil or whether it contained higher order structure. The interaction of this domain (or not) with the surface of the oil body will depend upon its secondary structure and, in particular, on the presence or absence of amphipathic regions within it.

In many polypeptides containing structurally distinct domains, these separate motifs fold to form independent but linked units of the protein. This enables the structure of the individual domains to be studied in isolation (15, 16). This observation is important since only relatively small protein domains are currently accessible to study by advanced techniques such as multidimensional NMR spectroscopy, which enable their structure at high resolution to be determined (15, 16). The drawback is that NMR spectroscopy is severely restricted for the study of some membrane proteins following reconstruction because of the non-isotopic motion of the proteins within the lipid bilayer matrix. This problem could be overcome by performing the study with small membrane-associated polypeptide domains and by studying the proteins in membrane-mimetic environments, such as in micelles and organic solvents. Another powerful strategy to surmount this problem is the use of physical techniques, such as CD and FTIR spectroscopy, for the analysis of full-length proteins or domains derived therefrom for the finer assignment and localization of secondary structures. Because domains are so much smaller than the full-length proteins, their secondary structures are much easier to assign by combining prediction data from the primary sequence with data from physical analysis.

According to our previous studies, the oleosin protein can therefore be considered as a protein of three domains, the central hydrophobic domain, the amphipathic C-terminal domain, and the polar N-terminal domain, each of which can be studied independently. In the present study, we report the expression of the N-terminal domain of the 18-kDa oleosin isolated from sunflower (12) as a recombinant fusion protein in Escherichia coli and the determination of its secondary structure using CD and FTIR spectroscopy either as a purified N-terminal domain in aqueous or organic solvents or as a liposome-reconstituted N-terminal domain. The structures derived from these physical studies were then compared with those predicted from the analysis of the primary sequence of the N-terminal domain.

**EXPERIMENTAL PROCEDURES**

Materials—Glutathione-Sepharose 4B, Sephadex G-75, and plasmid pGEX-3X were purchased from Sigma except Factor Xa, which was purchased from Boehringer Mannheim. All restriction and modification enzymes were purchased from Life Technologies. pfu DNA polymerase was purchased from Promega Corp. All other chemicals were of analytical grade.

DNA Manipulations—All DNA manipulations were performed by standard procedures (17). DNA sequencing was performed using the dideoxy chain termination method on double stranded DNA using Sequenase version 2.0 (U. S. Biochemical Corp.). E. coli strain DH5α was used for all DNA manipulations. E. coli strain JM109 was used for expressing the recombinant protein, and pGEX-3X vector (Pharmacia LKB Biotechnology Inc.) was used as an expression vector. pGEX-3X contains an open reading frame encoding glutathione S-transferase followed by unique restriction endonuclease sites for BamHI, SmaI, and EcoRI and termination codons in all forward reading frames. A schematic representation of the strategy for constructing plasmid pGEX-N is shown in Fig. 1. Using 10 ng of the amplified Sunflower cDNA as a template, digested with EcoRI (5' GTAC...5'GTAC...) and LMS (5'-GATC...) in the polymerase chain reaction under standard conditions (18), a 155-base pair fragment of DNA encoding the N-terminal domain of sunflower oleosin was amplified. Using a restriction site engineered into the LMS and a restriction site in the amplified polylinker this fragment was ligated into the BamHI/SmaI sites of expression vector pGEX-3X to generate pGEX-N. Plasmid DNA was prepared from several transformants and checked by restriction digest mapping and DNA sequencing. The predicted protein product of this in-frame fusion is a glutathione S-transferase-oleosin/N-terminal domain fusion. The N-terminal region can be cleaved from glutathione S-transferase via a Factor Xa cleavable sequence (IGYR). This plasmid was then used to transform JM109.

Expression in E. coli—Several transformants were grown individually in liquid culture in the presence of isopropyl thiogalactoside (IPTG). These transformants were screened for expression of expected hybrid protein by SDS-polyacrylamide gel electrophoresis (PAGE) of total cell lysate and for the ability of the soluble hybrid protein to bind to glutathione-agarose or to exhibit glutathione transference activity. A 10-ml overnight culture of a suitable transformant was then grown. This culture was diluted with 1 liter of LB + Kanamycin (10 μg/ml) and incubated (100 rpm) at 37°C. When the culture reached A650 = 0.5–1.0 (about 3 h), IPTG was added to 0.1 mM final concentration, and growth was continued at 37°C. After 3–7 h, cells were harvested by centrifugation (10 min, 3900 rpm), and resuspended in 10 ml of ice-cold PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM NaHPO4, 1.76 mM Na2HPO4, pH 7.4) with 2% Triton X-100.

Purification of N-terminal Polypeptide—Tubes containing the cells were immersed in ice, and the cells were lysed by using a sonicator with a 3-mm-diameter probe. The lysate was microcentrifuged for 5 min, and the supernatant was transferred to fresh tubes. The supernatant was applied to a glutathione-agarose column, and this was washed with PBS until no protein was eluted. Fusion protein was eluted with 50 mM Tris-HCl, 5 mM reduced glutathione (pH 8.0) buffer. Fractions eluted from the affinity column were collected and either analyzed by 12% SDS-PAGE or stored in aliquots at −70°C. Upon SDS-PAGE analysis, most of the hybrid protein was found to be contained in the insoluble phase (incubation bodies), with only a small amount in the supernatant. The inclusion bodies were strongly resistant to solubilization by mild detergent, so preparative electrophoresis was used to purify and denature the hybrid protein according to Gething and Von Jagow (19). An equal volume of 2 × SDS/sample buffer was added to the resuspended whole cell or cell lysate fraction. Samples were then boiled for 5 min, vortexed briefly, and loaded onto a 12% SDS-polyacrylamide gel. Gels were run for the appropriate time and removed into a clean tray, avoiding contact with fingers. Gels were rinsed with dideoxy HzO and stained with 5 μl of Coomasie Blue R-250. Gels were rinsed again and destained, if necessary, for 0–60 min with 50% methanol. Gel pieces containing the hybrid protein were cut out and placed into dialysis tubes in cathode buffer, and the hybrid protein was electroeluted. The eluted hybrid protein was dialyzed against 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM CaCl2, and 1 mM EDTA for 72 h. Then Factor Xa was added to the protein solution at 1:100 Factor Xa/Protein by weight. Incubation was carried out at 15°C for 20 h, and the efficiency of proteolytic cleavage was checked with SDS-PAGE. The mixture of uncleaved hybrid protein, carrier protein, cleaved N-terminal protein and Factor Xa was concentrated and separated by SDS-PAGE. The gel pieces containing the N-terminal peptide were cut out and put into a 10-ml siliconized tube. Gel pieces were soaked in two changes of dioxane H2O and 1 mM dithiothreitol for 15 min. The liquid was decanted and discarded, 5 ml of elution buffer (50 mM Tris-HCl, 0.1 mM EDTA, 5 mM dithiothreitol, 125 μM S'-Bicine, 0.2 M NaCl, and 0.1 M SDS) and the gel was crushed with 4–6 strokes of a small Teflon pestle that fitted snugly inside the test tube. The sample was then frozen and thawed several times. The N-terminal polypeptide was purified from the gel eluate using a YM 2 ultrafiltration membrane (1 kDa) and checked by SDS-PAGE.

Renaturation of N-terminal Polypeptide—Four volumes of cold acetone (20%) were added to the gel eluate, and the samples were allowed to precipitate for 30 min at −70°C. The tubes were centri-
N-terminal Domain of Oil Body Protein

bound proteins were visualized with Coomassie Blue stain, and the was replaced with polyvinylidene difluoride membrane for sequencing by semidry electrophoresis. The purified N-terminal polypeptide was concentrated, and the buffer in PBS (17) with vigorous mixing until a clear solution was obtained. The suspension was diluted into 100 ml containing 50 mM MgCl₂, 2 mM ATP, 10 mM glucose, and 5 µM GroEL and incubated at 25 °C for 6 h (21-23). GroEL was purified from E. coli as described (24). GroEL and bovine serum albumin were removed by ultrafiltration with the 10-kDa size cutoff (YM10) ultrafiltration Diaflo membrane, which allows the N-terminal domain to pass through the membrane while retaining GroEL and bovine serum albumin or alternatively by gel filtration column chromatography. The purified N-terminal polypeptide was concentrated, and the buffer was replaced with 30 mM phosphate buffer, pH 8.0, following ultrafiltration with 1-kDa size cutoff (YM2) ultrafiltration membrane. The purity was checked by SDS-PAGE.

Reconstitution of N-terminal Domain into Liposomes—L-α-Phosphatidycholine dioleoyl (PC) (3) was dried under nitrogen, and solvent-free PC was solubilized in a small volume (0.5 ml) of sodium cholate in PBS (17) with vigorous mixing until a clear solution was obtained with a final concentration of 20 mg/ml. The quantity of phospholipids dried down was calculated to produce a lipid to N-terminal protein weight ratio of 5:1. To dissolve the phospholipid, a greater than 4-fold excess (w/w) of sodium cholate was employed. The purified N-terminal domain (4 mg/ml) was then added to the lipid/detergent mixture, and the concentration of cholate was confirmed to be above its critical micellar concentration (7 mM). The mixture was centrifuged on a 520% sucrose gradient (6 h, 35,000 × g) at room temperature using a Bio-Rad FTS60 spectrometer and using a liquid nitrogen-cooled mercury-cadmium telluride detector at a resolution of 2 cm⁻¹. The samples were presented to the spectrometer in solution (30 mM Tris-HCl buffer, pH 8.0, at a concentration of about 2 mg/ml). A total of 1200 scans was collected for each spectrum. The control samples were 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. (27). In preparation for curve fitting, band contours were confirmed and second derivative spectroscopy. Secondary structure components were accomplished by least-squares iteration. Gaussian band shapes were assumed for the deconvoluted components.

RESULTS

Generation of E. coli Expressing the N-terminal Domain—A cDNA encoding the N-terminal domain 52 residues of the sunflower 18-kDa oleosin (12) was cloned into the pGEX-3X vector and placed under the control of the IPTG-inducible tac promoter (Fig. 1). E. coli JM109 cells were transformed with the pGEX-N plasmid, and colonies were picked. The DNA inserts were sequenced, and recombinant protein production was analyzed by SDS-PAGE. The size of the putative hybrid protein of 26 kDa, as measured by SDS-PAGE, was also consistent with a glutathione S-transferase-oleosin N-terminal fusion product (Fig. 2). The clone expressing the highest levels of hybrid protein was used for all subsequent studies.

Expression and Purification of Hybrid Protein—Induction of the tac promoter with IPTG in cells transformed with pGEX-N resulted in the synthesis of a polypeptide of about 26 kDa. In the absence of IPTG, the plasmid-encoded lacₐ allele is efficient in repressing transcription from the tac promoter. After induction of the tac promoter with IPTG, synthesis of an abundant glutathione S-transferase-oleosin N-terminal hybrid protein was observed on 12% SDS-PAGE. The harvested cells were lysed by using a probe sonicator, and extracts were separated into supernatant and pellet fractions, which were analyzed on 12% SDS-PAGE (Fig. 2). A protein of approximately 26 kDa was observed only in the cells containing the pGEX-N plasmid and only in the presence of IPTG (Fig. 2). The protein was primarily in the insoluble fraction and represented 30–50% of insoluble protein based on Coomassie Blue staining. However, as verified by glutathione-agarose affinity chromatography of the cell supernatant fraction on immobilized glutathione only a small amount of hybrid protein existed in the aqueous phase. The hybrid protein in the insoluble fraction (inclusion body) could not be solubilized by mild detergents nor could the purified hybrid protein in its native form be cleaved by Factor Xa from the affinity column. Rather, it was necessary to denature the purified hybrid protein in the presence of SDS before it was possible to cleave it using Factor Xa. Although the hybrid protein prepared by affinity chromatography was over 90% pure on SDS-PAGE, it was found to be easily degraded, possibly due to trace amounts of proteases in the preparation.
**N-terminal Domain of Oil Body Protein**

**Fig. 1. Strategy for constructing plasmid pGEX-N.** Using oligonucleotide primers RP and LM3 and plasmid EE as a template, the open reading frame (ORF) encoding the N-terminal domain of the sunflower oleosin was amplified using the polymerase chain reaction (PCR). Gel purification of this product and further digestion by EcoRI and BamHI allowed it to be cloned into the BamHI/EcoRI restriction sites of pGEX-3X to make pGEX-N. The recombinant plasmid encodes a glutathione S-transferase (GST)-oleosin N-terminal fusion, the expression of which is under the control of the Ptac promoter.

**Fig. 2. SDS-PAGE analysis of expressed and purified N-terminal domain.** Lane 1, molecular mass markers; lane 2, non-recombinant JM109 whole cell lysate; lane 3, recombinant JM109 whole cell lysate; lane 4, extracellular protein; lane 5, affinity-purified fusion protein; lane 6, SDS-PAGE-purified fusion protein; lane 7, the fusion protein digested with Factor Xa (fusion protein, glutathione S-transferase, and the N-terminal polypeptide can be detected in this lane); lane 8, carrier protein purified by affinity chromatography of the digested fusion protein; lane 9, purified N-terminal polypeptide; lanes 1-9 are Coomassie Blue-stained; lane 10, Factor Xa-digested fusion protein, silver stained.

Therefore, the purification method was changed to preparative electrophoresis to obtain sufficient quantities of stable, denatured, Factor Xa cleavable hybrid protein directly from inclusion bodies.

**Purification of N-terminal Domain**—The utility of the pGEX-3X vector for production of foreign polypeptides in *E. coli* would be increased if the glutathione S-transferase carrier protein could be removed completely from the hybrid protein by cleavage with a site-specific protease. This cleavage has sometimes been ineffective due to insolvency or the presence of denaturing reagents (39). In the present case, lack of cleavage was probably caused by the inaccessibility of the Factor Xa site in the native hybrid protein. Purified denatured hybrid protein was cleaved following incubation of the protein in a solution containing Factor Xa for 24 h at 15 °C, and the mixture containing hybrid protein, carrier protein (glutathi-one S-transferase), and N-terminal protein was then separated by SDS-PAGE gels (Fig. 2). The N-terminal domain was isolated from other proteins by SDS-PAGE after Factor Xa cleavage. The N-terminal protein was eluted from gels by several freeze-thaw cycles in eluting buffer, separated and concentrated by ultrafiltration, and checked on 15% SDS-PAGE (Fig. 2). It was stained by double staining of Coomassie Blue and silver staining, because the N-terminal domain was not easily stained by Coomassie Blue only.

N-terminal sequence analysis of the purified polypeptide revealed that the first 10 amino acid sequences matched the sequence predicted from the cDNA sequence of the sunflower oleosin, i.e. GIRVPWTTTT (12).

**Renaturation of N-terminal Polypeptide in the Presence of GroEL**—The conventional method of renaturing polypeptides purified from SDS-PAGE gels (40), *i.e.* the progressive removal of denaturant by dialysis, was not successful as shown later in Fig. 5, where no constant secondary structure can be detected upon CD analysis. A possible explanation for this is that most of the eluted N-terminal polypeptide was denatured into a very flexible form. We examined this hypothesis by renaturing the N-terminal polypeptide in the presence of the GroEL purified from *E. coli*. GroEL is a heat-shock protein that acts as a chaperonin and promotes the correct folding of denatured proteins in *vitro* (41, 42). Since chaperonins such as GroEL bind only unfolded proteins, not those that are fully folded, the N-terminal protein was denatured completely before refolding to avoid misfolding. All denaturants were then diluted 100 times in a solution containing GroEL. Upon adding ATP, GroEL apparently bound the denatured N-terminal polypeptide causing the latter to adopt a constant secondary structure and then be released from GroEL as checked by CD spectroscopy (Fig. 5). We therefore confirmed that the N-terminal polypeptide was completely denatured following SDS purification. The denatured polypeptide could not easily be refolded by conventional methods (21) but could refold in the presence of GroEL.

**Reconstitution of N-terminal Domain into Liposomes**—After 30 h of dialysis the N-terminal domain-free dialysis control
had become cloudy, but the reconstitution complex was much more transparent. After formation of the liposomes by dialysis, the remaining free N-terminal domain protein was removed from the N-terminal domain bound to liposomes by centrifugation to sediment the well formed pellet of the protein-liposome complex. This pellet was washed in dialysis buffer and resuspended in 10% sucrose in Tris-HCl buffer. The suspension was fractionated by centrifugation on a sucrose density gradient. A band of density 1.02 g/ml representing liposomes containing little or no protein was formed near the top of the gradient. The protein-containing liposomes were shown to be at 1.05 g/ml. The presence of the N-terminal domain was checked by SDS-PAGE electrophoresis. The concentration of N-terminal protein was determined at this stage.

Secondary Structure Prediction from Primary Sequence Data—The N-terminal domains of oleosins from different plants are highly variable, and they all have hydrophobic and hydrophilic residues distributed through the polypeptide. In the present case, no definite specific secondary structure can be assigned, although it is likely that the sunflower N-terminal domain contains less than 10% α-helical structure-forming potential, i.e. from residues 11-17 and 24-30, as shown in Fig. 3. Some β-strand structures are also potentially formed near the N and C termini of this polypeptide from residues 2-10 and 42-48 (Fig. 3). If a helical wheel is drawn in the putative α-helical forming sequence, an amphipathic α-helical structure can be identified between residues 11 and 22 as shown in Fig. 4.

Estimation of the Secondary Structure Content by CD—The CD spectra of the N-terminal polypeptide with different treatments are plotted together in Fig. 5 for ease of comparison. The general features of the spectra were different although they are derived from protein samples that have the same sequence and concentration. When the N-terminal domain was refolded using the conventional methods (21) no definite secondary structure was observed (Fig. 5B). If the denatured N-terminal domain was solubilized in 50% TFE to mimic the membrane environment, a secondary structure was displayed. Analysis showed that in this environment, the polypeptide contained approximately 59.6% β-structure, 40.4% unordered structure, and no definite α-helical structure (Fig. 5B). This is consistent with the finding of increased amounts of α- and β-structures in full-length oleosin proteins, solubilized in 50% TFE solvent (data not shown). It is likely that the secondary structure was affected by lipid association. The β-structure could be further subdivided into approximately 10% β-strand and 40% β-turn by the Chen et al. method (34, 35). After refolding in the presence of the chaperonin GroEL, the denatured polypeptide adopted a constant secondary structure and consisted of approximately 10% α-helical structure, 28% β-structure, and 62% unordered structure (Fig. 5A). The α-structure content could be further subdivided into approximately 21% β-strand and 8% β-turn structure by the Chen et al. method (34, 35). This result confirmed that GroEL binds only the unfolded proteins and releases them in a folded form. In contrast, when the N-terminal polypeptide was reconstituted into liposomes, the content of secondary structure was changed and contained approximately 21% α-helical structure, 31.8% β-structure, and 47.2% unordered structure by the CONTIN program (Fig. 5A) and 20% α-helical structure, 26% β-structure, 14% β-turn structure, and 40% unordered structure by the method of Chen et al. (34, 35). After reconstitution of N-terminal domain into liposomes, the content of α-helical structure and β-strand structure was increased sig-

![Fig. 3](image-url)
TABLE I
Fractional secondary structure content of the N-terminal domain of an oleosin derived from CD spectra data

<table>
<thead>
<tr>
<th></th>
<th>α-Helix</th>
<th>β-Strand</th>
<th>β-Turn</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDa</td>
<td>CDb</td>
<td>CDc</td>
<td>CDd</td>
<td></td>
</tr>
<tr>
<td>Denatured</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>50% TFE</td>
<td>ND</td>
<td>59.6</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>Renatured</td>
<td>12</td>
<td>28</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>Reconstituted</td>
<td>21</td>
<td>20</td>
<td>31.8</td>
<td>26</td>
</tr>
</tbody>
</table>

* Analysis by the method of Provencher and Glockner (33).
* Analysis by the method of Chen et al. (34, 35).

The percentage of β-structure means the combined value of β-sheet and β-turn by this method. ND, no secondary structure detected.

Conformational Assignment of Reconstituted N-terminal Domain with Liposome from FTIR Spectrum—A typical FTIR spectrum of the N-terminal domain reconstituted into liposomes in solution is shown in Fig. 6. Fig. 6 shows the deconvolution FTIR spectrum of the amide I, II, and III regions, ranging from 1000 to 1800 cm⁻¹. The frequencies of the main peaks and their tentative assignments are listed in Table II. The main amide I peaks of the reconstituted N-terminal domain occur at 1672.3 cm⁻¹ with shoulders at 1659.4 and 1649.7 cm⁻¹ together with two small peaks at 1629.7 and 1616.3 cm⁻¹. The main amide II peaks are found at 1547.5, 1579.7 cm⁻¹ with shoulders at 1563.2, 1529.8 cm⁻¹. The amide III peaks occur at 1245, 1262 cm⁻¹ with shoulders at 1228.5 cm⁻¹.

In IR spectra, the amide bands that arise from the vibration of the peptide groups provide information on the secondary structure of polypeptides and proteins. Both theoretical and experimental studies with model polypeptides have shown that there is good correlation between the amide I band...
frequency and the type of secondary structure present. Based on numerous observations on synthetic polypeptides and proteins (44-46), the broad peaks in the FTIR deconvoluted spectrum of the N-terminal domain at 1672.3, 1517.9, 1245, and 1228.5 cm⁻¹ are indicative of β-strand structure. The main peak at 1672.3 cm⁻¹ is indicative of parallel β-strand structure, and the peak at 1629.7 cm⁻¹ is indicative of antiparallel β-strand structure, which shows agreement with the CD data (26% β-strand structure) for the N-terminal domain reconstituted with liposomes. Peaks at 1649.7, 1616.3, 1547.5 cm⁻¹ are indicative of α-helical structure in the polypeptide, which also corresponds closely with the CD data (20% α-helical structure) for the reconstituted N-terminal domain with liposomes. Peaks at 1659.4 and 1262 cm⁻¹ are indicative of β-turn structure. However, it is also suggested that the broad peak at 1672.3 cm⁻¹ might be indicative of β-strand or mixed domains (44-46). It is quite difficult to assign a particular subregion of amide I, II, and III regions to unordered structures, because there is normally overlap of α-helical absorption with that of random coil. In qualitative terms, the results of the FTIR analysis of the reconstituted N-terminal domain agree well with the data from CD analysis, i.e. in showing a large amount of β-structure and a smaller but still significant amount of α-structure.

**DISCUSSION**

Previous studies in our laboratory (11) have predicted the structure of the major 19-kDa oleosins from rapeseed. Based on these predictions, the oleosin molecule can be divided into three separate structure domains: (a) a relatively polar N-terminal domain of 53 residues containing a predicted 18-residue α-helical region; (b) a central hydrophobic β-strand domain of about 75 residues; and (c) a polar C-terminal domain of about 65 residues containing a potential amphipathic α-helical stretch of 20-25 residues exhibiting similarities with the 11-residue amphipathic α-helical repeats found in most animal apolipoproteins (10, 11). Further physical studies (11) have confirmed that oleosins contain about 45% β-strand and 13% α-helical structure. To resolve the structure of the oleosin molecule in finer detail, it is desirable to express the three domains of the protein and examine each in detail. This strategy serves three purposes. First, it allows us to verify that the oleosin molecule is indeed made up of independently folding domains, as reported for several other classes of protein (14-16). Second, the use of smaller protein domains facilitates the investigation of the secondary structures at finer resolution using physical techniques, i.e. the secondary structure can be assigned to a domain level and can greatly improve the full protein model in terms of detailed structural assignments. What is more important is that, due to the non-isotropic motion of membrane proteins within the lipid matrix, NMR spectroscopy is severely restricted for the study of membrane protein in the reconstituted condition. Hence the secondary structure under reconstituted conditions could only be determined by other physical techniques, such as CD and FTIR. Third, the expression of relatively small (<10 kDa) domains allows the future study of their fine structure by two-dimensional NMR spectroscopy in some membrane-mimetic environments, such as in micelles and organic solvents at resolutions hitherto achievable only by x-ray crystallography (15, 16). To this end, we have expressed and characterized the N-terminal domain corresponding to residues 1-52 of the 18-kDa oleosin from sunflower.

Using a polymerase chain reaction-based strategy, DNA encoding the N-terminal domain was cloned into the pGEX-3X expression system, and *E. coli* cells were transformed with this construct. Yields of the fusion protein were high. Although purification of this fusion protein was efficient, cleavage of the protein at the Factor Xa cleavage site was rather inefficient. This inefficiency was probably due to the obstruction of the target site by structural features in the immediate vicinity of the site. To circumvent this problem, the fusion protein was first fully denatured in 6 M guanidine HCl, diluted, cleaved with Factor Xa, and the Factor Xa was removed. The cleaved domain was then renatured in the presence of the chaperonin GroEL, and the latter was removed from the well folded N-terminal domain. This renaturing step caused the protein to have a stable secondary structure, as proved when these samples were subjected to CD spectroscopy.

Sodium cholate, which is technically simple to remove to unperturbing levels, was then used to form protein-lipid complexes in PC liposomes through detergent dialysis. After 150 h of dialysis, the protein-free control was cloudy, whereas the protein-lipid complex was quite transparent. After removing lipid-free protein and liposomes containing little or no protein by centrifugation and sucrose density gradient centrifugation, the solution of the protein-lipid complex was still quite transparent and stable. It is suggested that the N-terminal protein has emulsifying effects on the liposomes as an interfacial emulsifying agent.

Circular dichroism spectroscopy has now come of age. With modern commercial CD instrumentation that makes measurements far into the UV region, it is possible to obtain data with a much higher information content than was possible hitherto. Recently published methods for analyzing CD spectra are fairly reliable, so it is now possible to monitor the secondary structure of proteins in solution. It is well documented that CD is an excellent method for the determination of the α-helical content of proteins with a high degree of accuracy. However, the estimation of β-structures by CD is less satisfactory. In contrast, FTIR spectra of proteins can provide accurate estimation of β-structure in protein. Common features of IR spectra of protein are the so-called “amide” bands, which arise from delocalized vibrations of the peptide linkage. Of these, the amide I band (1700–1600 cm⁻¹) is the most useful for the analysis of protein secondary structure. Correlations of amide I band frequency with the presence of α-helical, antiparallel and parallel β-strand, and random coil structures in protein in H₂O and D₂O are well established (44). Unfortunately, this technique is less effective than CD in the determination of α-helix content. For these reasons, both techniques were used in the present study to obtain as much information as possible about the secondary structure of the N-terminal oleosin domain reconstituted into liposomes.

Our previous structure predictions from the primary sequence of the rapeseed oleosin N-terminal domain indicated some secondary structure although definite assignments were not possible (11). In contrast, predictions from the maize oleosin N-terminal domain indicated very little defined secondary structure (14). In the present study, it is shown that the purified, renatured, and reconstituted sunflower N-terminal domain consisted of about 10-20% α-helical and 30-40% β-strand structure with approximately 40-50% of the polypeptide made up of random coil structure. Due to the considerable divergences in amino acid sequence in the N-terminal domains of oleosins from different plants (13), it is possible that they do indeed adopt very different secondary structures. Alternatively, when these domains are interacting with a lipid:water interface on the surface of an oil body in *vivo*, it is possible that the additional constraints imposed by
this environment force them to adopt more similar structures. The resolution of this question can only be achieved by studying the secondary structures of purified oleosin domains in environments close to the in vivo situation. The surface of an oil body, where the N-terminal domain is located, is made up of a monolayer of phospholipid enclosing a matrix of triacylglycerol (10–12). Due to their size and high content of polar residues, it is unlikely that the N-terminal domains of any oleosins identified to date can extend deeply into the matrix of the oil body. Rather, they probably interact with the phospholipid monolayer alone. Therefore the liposome reconstitution system, which is technically much easier to prepare and analyze by CD and FTIR than reconstituted oil bodies, was used in the present study as a close mimic of the in vivo system.

The data, derived from FTIR spectroscopic studies, showed the peptide is well structured with some antiparallel and some parallel β-strand structures plus α-helical and β-turn structures. The areas of different structural components are consistent with the result derived from CD data by the method of a least-square curve-fitting program. It is predicted that residues 10–23 and 43–49 would produce an α-helical structure. It is likely that some of the 10–25% α-helical structure exists in these areas. From a helical wheel drawn with residues 10–23, we see that this putative α-helix is amphipathic. Residues 30–37 and 42–48 would produce a β-strand structure at the N and C terminus of this polypeptide, and residues 2–9 may produce an antiparallel β-strand structure. A model for this structure is presented in Fig. 7.

Our preliminary expression experiments confirmed that free oleosins are very labile in the E. coli expression system. Fusing them, even as small domains, with glutathione S-transferase significantly stabilizes them or their domains, and they could be accumulated to high levels in the cell. Although Factor Xa is considered a highly specific protease, we observed they could be accumulated to high levels in the cell. Although functional studies on proteins are often relatively easy, structural studies continue to be the rate-limiting step in their analysis. Structural biologists are therefore constantly searching for new techniques or improvements of existing methods to increase the pace of protein structural determinations. The complete three-dimensional structure of a protein at high resolution can be determined by x-ray crystallography or, as an alternative to diffraction, multidimensional NMR spectroscopy. However, each of these techniques has limitations and is not possible to employ for all proteins. Oleosins are unusual very hydrophobic proteins containing a central nonpolar domain flanked by relatively polar C- and N-terminal domains (10–14). Because of this, oleosins are very difficult to crystallize, even when purified in native form. Due to their low solubility in both aqueous and non-aqueous solutions and non-isotropic motion of the proteins within the lipid matrix, it is also very difficult to study oleosins by NMR. For these reasons, we have chosen to undertake structural studies of isolated oleosin domains. The three oleosin domains identified in this study, i.e. N-terminal, central, and C-terminal, are quite distinct from one another both in structure and location. While the central domain is likely to be fully immersed in the triacylglycerol matrix of the oil body, the N- and C-terminal domains probably reside at the phospholipid-water interface (11). Therefore, it is justified to study each of the three domains in isolation.

In summary, the 53-residue N-terminal domain of an oleosin protein from sunflower has been expressed, purified, and reconstituted in various systems, including liposomes. In the latter system, which is closest to the in vivo situation, the domain contained considerable secondary structure made up of about 10–20% α-helix and 30–40% β-structure. Part of the putative α-helical structure was amphipathic, indicating that it could extend across the lipid-water interface of the oil body in vivo. Such amphipathic α-helices provide recognition sites for lipase binding in apolipoprotein C-II, which also resides at the surface of a triacylglycerol matrix bounded by polar lipids, i.e. the chylomicron (10). Whether or not the N-terminal domain of oleosins plays a similar role in lipase binding during postgerminative lipid mobilization in seeds remains to be determined.

Acknowledgments—We thank all members of the oleosin research group for their advice and assistance.

REFERENCES
N-terminal Domain of Oil Body Protein