PROTEINS: Structure, Function, and Genetics 38:341–349 (2000)

# Structural Characterization of a Methionine-Rich, Emulsifying Protein From Sunflower Seed

**Maya J. Pandya,[1,2]* Richard B. Sessions,[1] Phil B. Williams,[1] Christopher E. Dempsey,[1] Arthur S. Tatham,[2] Peter R. Shewry,[2] and Anthony R. Clarke[1]**

[1]*Molecular Recognition Centre and Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol, United Kingdom*

[2]*IACR - Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Bristol, United Kingdom*

***ABSTRACT*** **The 2 S seed storage protein, sunflower albumin 8, contains an unusually high proportion of hydrophobic residues including 16 methionines in a mature protein of 103 amino acids. A structural model, based on the known structure of a related protein, has been constructed as a four-helix bundle cross-linked by four disulphide bonds. This model structure is consistent with data from circular dichroism and nuclear magnetic resonance experiments. Analysis of the model's surface shows the presence of a large hydrophobic face that may be responsible for the highly stable emulsions this protein is known to form with oil/water mixtures. Proteins 2000;38:341–349.** © 2000 Wiley-Liss, Inc.

**Key words: seed storage protein; methionine-rich; emulsifying agent; homology model; hydrophobic face**

## INTRODUCTION

Developing seeds of plants synthesize and accumulate proteins whose major function is to provide a store of nitrogen, sulphur, and carbon. The 2 S albumins of sunflower (*Helianthus annuus* L.) account for about a third of the total salt-soluble proteins, the remainder being predominantly 11S globulins (helianthinins).[1] The major methionine-rich sunflower albumin, called SFA-8, is known to form highly stable emulsions with oil/water mixtures,[2] implying the presence of a large exposed surface of hydrophobic residues.[3,4] SFA-8 contains an unusually high proportion of hydrophobic residues including 16 methionines in a mature protein of 103 amino acids.[5] The protein also contains four disulphide bonds[6] and shows no tryptic, chymotryptic, or α-amylase inhibitory activity in vitro, despite the presence of a putative trypsin inhibitory site.[6] Although 2 S albumin proteins have been characterized from a wide range of species,[7] little detailed structural information and no three-dimensional structures have been available until recently. A preliminary nuclear magnetic resonance (NMR) structure has been determined for a 2 S albumin from rape seed, which is a five α-helix bundle consisting of two polypeptide chains.[8]

Here, we propose a model for the structure of the protein based on the known structure of a related protein, a bifunctional α-amylase/trypsin inhibitor from seed of finger millet[9] (RBI), which is substantiated by the results from dichroism (CD) and NMR spectroscopy. The model structure consists of a bundle of four α-helices cross-linked by the four disulphide bonds, and supports the idea that the molecule has a large hydrophobic face.

## MATERIALS AND METHODS

### Homology Modeling

#### *Identifying a template structure*

The amino acid sequence of SFA-8[5] was compared with the sequences of proteins of known structure by a BLAST search in the Structural Classification of Proteins (SCOP) database.[10] Three hits with significant E values over part of the sequence occurred, 1BIP/1JFO (E 0.00030, 35 residues), 1PNB (E 0.00012, 20 residues), and 1HSS (E 0.015, 22 residues). All three of these proteins belong to a single-fold family described in SCOP as "4 helices; folded leaf; right-handed superhelix; disulphide-rich." This family currently contains seven members, and the structures of these are represented in Figure 1. Pairwise sequence alignments of SFA-8 with these seven proteins showed low sequence identities (Table I) barely above what might be expected from matching a random sequence. Based on this information alone, assignment of SFA-8 to this fold would be tenuous; however, the family is also characterized by four conserved disulphide linkages. The appropriate cysteine residues are also present in SFA-8, and experiment has shown these to have an appropriate connectivity[6] (Fig. 6). A multiple sequence alignment using Clustal is shown in Figure 3a, and graphically illustrates the difficulty of sequence alignment in circumstances of low sequence homology, with only half of the cysteine positions perfectly aligned. Nevertheless, the pattern of cysteine residues strongly supports the assignment of SFA-8 to this fold family. The best candidates for the homology modeling structural template are the bifunctional inhibitor from

*Abbreviations:* CD, circular dichroism; CHAPS, (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane-sulfonate); COSY, correlation spectroscopy; DQF, double quantum-filtered; DTT, dithiothreitol; GdmCl, guanidinium chloride; GdmSCN, guanidinium thiocyanate; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; HPLC, high-performance liquid chromatography; NOESY, nuclear Overhauser spectroscopy; PAGE, polyacrylamide gel electrophoresis; pH*, reading in $D_2O$ using a hydrogen electrode; PMSF, phenylmethylsulfonylfluoride; TFE, trifluoroethanol.

*Correspondence to: M.J. Pandya, Department of Biochemistry, School of Biological Sciences, University of Sussex, Falmer, East Sussex, BN1 9QG, England. E-mail: maya@biols.sussex.ac.uk

Received 4 June 1999; Accepted 6 September 1999

ragi (Indian finger millet)[9] (pdb codes 1BIP, 1JFO now 1B1U), referred to here as RBI, and the napin seed storage protein (pdb code 1PNB),[8] but the latter is a poor candidate because the polypeptide is cleaved between residues Phe 31 and Gln 32 and these residues are >20 Å apart in the structure. Both NMR and X-ray structures of RBI are available, and are very similar, as illustrated in Figure 1 (Cα root mean square = 2.1 Å for residues 6–112), hence RBI and the sequence alignment shown in Figure 3b were used as the basis of the SFA-8 homology model. A secondary structure prediction using the Jpred[11] server at EBI was carried out on the SFA-8 sequence, yielding the consensus result shown in Figure 2a.

### *Building the model*

The nine residues at the N-terminus of SFA-8 were omitted from the structural model because there is no homology basis for modeling these residues, nor does the sequence of these amino acids suggest any secondary structure (Fig. 2a). The 12 residues from Tyr 12 to His 23 in SFA-8, which are predicted to be helical, were introduced as a stretch of regular α-helix. This range of residues is not helical in RBI, but occurs as helix in the lipid transfer proteins (1JTB, 1LPT, 1MZM) and is partly helical in the napin albumin 1PNB. The three-residue insertion between Cys 29 and Gly 30 of the RBI template was achieved by replacing the four-residue stretch Gly 30 to Pro 33 with a seven-residue loop (Thr 56 to Pro 62) from horse liver alcohol dehydrogenase (pdb code 8ADH), chosen by geometric matching from a library of known structures. The 16-residue deletion in SFA-8 corresponds to a β-hairpin in RBI (residues Asp 65–Gln 80). Some minor remodeling of the residues on either side of the deletion site allowed the backbone to be rejoined. The 14 residues from Arg 80 to Glu 93 of SFA-8 were introduced as a stretch of regular α-helix, because this region is strongly predicted to be α-helical (Fig. 2a). The final deletion corresponds to a four-residue reverse β-turn between Thr 107 and Gly 110 in RBI; this turn was excised and residues 108 and 112 joined. The C-terminal residues Leu 117 to Glu 122 in RBI were deleted. The remainder of the SFA-8 sequence was introduced by direct substitution according to the alignment in Figure 3b.

Side-chain clashes in the resultant model were relieved by altering torsion angles by inspection. The structure was covered in a 5-Å layer of water and energy minimized until the average absolute derivative of coordinates with respect to energy fell below 0.01 kcal $mol^{-1}$ $Å^{-1}$. Analysis of the minimized structure using Procheck[12] showed the model to possess acceptable geometry, as shown by the Ramachandran angles and G factors of Figure 2b. A short molecular dynamics simulation was carried out to ensure that the model structure is located in a minimum-energy well in the potential surface. The simulation was carried out in a box of water under periodic boundary conditions for 100 ps at 350 K, and structures were energy minimized every 10 ps. The Cα traces of these structures, superimposed on the initial model, are shown in Figure 7. Visualization, building, and manipulation of structures was carried out using InsightII 2.35 and energy calculations with Discover 2.95 (MSI) using the consistent valence forcefield. The model structure was analyzed for the presence of surface hydrophobic patches using the patch-delineating program Quilt.[13]

### Purification of SFA-8

2 S albumins were purified from sunflower seed (*Helianthus annuus* L, cv Alphasol, Twyfords Seeds UK Ltd.) by a modification of the method of Kortt and Caldwell.[14] The seed was mechanically dehulled and the kernels were ground to a flour in a blender. The flour was defatted twice by stirring with petroleum spirit (60–80° bp) (10:1 vol/wt, solvent:flour) for 1 h at 4°C, and the solvent was decanted. The defatted flour was air dried and then extracted with 20 mmol/L Tris/HCl, pH 7.8, containing 0.5 mol/L NaCl and 1 mmol/L PMSF (10:1, vol/wt, buffer:flour), by stirring for 2 h at 4°C. The extract was centrifuged (18,500 *g*, 30 min at 4°C) and the globulins were precipitated from the supernatant by adding cold methanol containing 1 mmol/L PMSF to give a final concentration of 60% (vol/vol) methanol. After centrifugation (12,000 *g*, 30 min at 4°C), the 2 S albumins were precipitated from the supernatant by the addition of 3 vol of acetone and standing at −20°C for 16 h. After decanting the solvent, the precipitated albumins were dissolved in and dialyzed against water using a membrane with an $M_r$ cutoff of 3,500 (Spectra/Por 3, Pierce and Warriner), and then freeze dried.

The total 2 S albumin fraction was separated by cation exchange chromatography on a column (1.6 × 25 cm) of S-Sepharose equilibrated in 50 mmol/L Na citrate buffer, pH 4.5. After washing the column with starting buffer to remove the majority of the 2 S albumins, a linear gradient of 0–0.5 mol/L NaCl over 500 ml was applied. The SFA-8-enriched fraction, which was the first of three protein peaks (detected by $A_{280}$) was dialyzed against water and freeze dried. SFA-8 was purified further by reversed-phase (RP)-HPLC on a Vydac 218TP1010 column of $C_{18}$ (1 × 25 cm; Sigma-Aldrich). After washing the column with 30% (vol/vol) acetonitrile in 0.05% (vol/vol) aqueous trifluoroacetic acid (TFA), the column was eluted with a 75-ml linear gradient of 30–45% (vol/vol) acetonitrile in 0.05% (vol/vol) aqueous TFA, and the SFA-8 peak (detected by $A_{225}$) was collected and freeze dried. The purity of SFA-8 samples was monitored by Tris/tricine sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (SDS-PAGE) in the presence of the reducing agent, 2-mercaptoethanol.[15] Two hundred grams of dehulled seed yielded about 80 mg of essentially pure SFA-8. Freeze-dried samples were dissolved when required for experiments, and concentrations were estimated by ultraviolet (UV) absorption at 280 nm ($\epsilon$ = 8,800 mol/$L^{-1}$ $cm^{-1}$).

### Mass Spectrometry

Mass spectra were run on a VG Quattro triple quadrupole instrument with electrospray ionization. The sample was introduced in 50% (vol/vol) acetonitrile/water, containing 1% formic acid, at a flow rate of 20 μl $min^{-1}$.

**TABLE I. Pairwise Clustal Sequence Alignments Between SFA-8 and the Seven Members of the SCOP "Four Helices; Folded Leaf; Right-Handed Superhelix; Disulphide-Rich" Fold Family**

| Protein | % Identity | % Strong similarity | % Weak similarity | % Different |
|---|---|---|---|---|
| RBI | 23.7 | 16.0 | 4.6 | 55.7 |
| 1HSS | 22.6 | 11.7 | 3.7 | 62.0 |
| 1PNB | 19.5 | 24.8 | 7.1 | 48.6 |
| 1HYP | 18.5 | 9.3 | 8.3 | 63.9 |
| 1LPT | 16.2 | 11.4 | 15.2 | 57.2 |
| 1JTB | 15.9 | 13.3 | 11.5 | 59.3 |
| 1MZM | 14.7 | 11.9 | 6.4 | 67.0 |

### Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge at 22,000 rpm and 20°C. Data were analyzed by fitting to the Lamm equation to obtain molecular weights:[16]

$$ln\ c_r/r^2 = (M(1 - v\rho)\omega^2)/(2RT)$$

where $c_r$ is the solute concentration at radial distance $r$, $M$ is the monomer molecular weight, $v$ is the partial specific volume, ρ is the solvent density, ω is the angular velocity (radians/second) , $R$ is the gas constant, and $T$ is the temperature (K). Samples of SFA-8 at 1.2 mg $ml^{-1}$, 0.5 mg $ml^{-1}$, and 0.2 mg $ml^{-1}$ in 50 mmol/L Tris/HCl buffer, pH 7, were analyzed. Further sedimentation equilibrium experiments were performed on samples of SFA-8 at higher concentrations in a Beckman Model E analytical ultracentrifuge at 12,000 rpm and 20°C.

The molecular weights given in the results are the z-average molecular weight:

$$M_z = (\Sigma\ N_i M_i^3)/(\Sigma\ N_i M_i^2)$$

which is the default average obtained from schlieren optics and where $N_i$ is the number of molecules of molecular weight $M_i$. Samples of SFA-8 at 23 mg $ml^{-1}$ in 50 mmol/L Tris/HCl buffer, pH 7, and at 31 mg $ml^{-1}$ in 50 mmol/L sodium acetate buffer, pH 4.5, were analyzed. Values of 1 g $ml^{-1}$ and 0.717 ml $g^{-1}$ were taken for the buffer density and the partial specific volume of the protein, respectively.

### CD Spectroscopy

Far-UV CD spectra (185–250 nm) were recorded in 0.1-mm pathlength cells with a Jobin Yvon CD6 spectropolarimeter. A sample of native SFA-8 at 0.15 mmol/L in 25 mmol/L K phosphate buffer, pH 7, was examined at 25°C. Data accumulations using 0.5-nm step resolution, 1 nm $min^{-1}$ scan rate, and 2-nm bandwidth are reported with baseline subtraction. The fractional α-helical content ($f_H$) of SFA-8 was estimated from the molar ellipticity measurement at 222 nm ($\theta_{222}$) using the following algorithm[17]:

$$\%\ \alpha\text{-}helix\ =\ (100(\ -\ \theta_{222}\ -\ 2340))/\ -\ 30300$$

### NMR Spectroscopy

NMR spectra at 27, 40, 50, and 70°C were obtained using a Jeol Alpha 500-MHz spectrometer. Spectra were acquired in phase-sensitive mode with quadrature detection and presaturation of the water signal using Dante pulse irradiation[18] during the relaxation decay. The following two-dimensional experiments were run: NOESY[19] with a mixing time of 180 ms; HOHAHA[20] with a mixing time of 150 ms; and DQF-COSY.[21] All two-dimensional spectra were obtained without spinning and with the irradiation and observation frequencies matched. The spectral width was set to ±3,000 Hz around the residual solvent signal. Data sets were collected with 2,048 real points in the $t_2$ dimension and 512 $t_1$ increments. These sets were zero-filled in $t_1$, and shifted sine bell or sine squared bell window functions were applied in both dimensions before Fourier transformation. The data were baseline-corrected after transformation in the $t_2$ frequency dimension, and $t_1$ ridge suppression[22] was used after transformation of NOESY and HOHAHA data. Data processing was performed using Felix 2.30 software (MSI).

SFA-8 was dissolved at a final concentration of 2–6 mmol/L in 50 mmol/L sodium acetate buffer, pH* 4.5, containing 10% $D_2O$ and 0.05% sodium azide. Additional spectra were collected for SFA-8 dissolved in buffer containing either 20% (vol/vol) TFE, 2 mol/L GdmCl, 1.5 mol/L GdmSCN, or 20 mmol/L CHAPS[23] to suppress protein self-association. pH readings in $D_2O$ are direct measurements using a hydrogen pH electrode, without correction.

### Proteolytic Digestion

A solution of native SFA-8 at 0.9 mg $ml^{-1}$ was preincubated for 15 min before digestion with trypsin at 1% (wt/wt) protease/substrate in 50 mmol/L Tris/HCl, pH 7.6, at 25°C. Aliquots were removed at 0, 0.5, 1, 2, 5, 10, 15, 30, 45, and 60 min, and digestion was terminated by adding SDS sample buffer and boiling. Control samples were prepared by omitting either trypsin or substrate. Samples were analyzed by SDS-PAGE in the presence of 2-mercaptoethanol[15] using low $M_r$ marker proteins (BDH). Protein bands from tryptic digest gels were transferred to ProBlott membrane (Applied Biosystems) and sequenced for 10 cycles of Edman degradation using an Applied Biosystems 477A protein sequencer.

## RESULTS AND DISCUSSION

### Homology Modeling of SFA-8

Although a number of 2 S albumin proteins have been characterized,[7] the only detailed structural information available is a preliminary NMR structure recently determined for a 2 S albumin from rapeseed, napin BnIb, which is a disulphide-linked five α-helix bundle consisting of two polypeptide chains.[8] Sunflower (*H. annuus* L.) seed contains at least 12 individual 2 S albumin proteins, which account for about 20% of the total seed proteins.[14,24] SFA-8 is synthesized as a precursor with an N-terminal hydrophobic prosequence and signal peptide. The mature protein is a monomer of 103 amino acids with four disulphide bonds

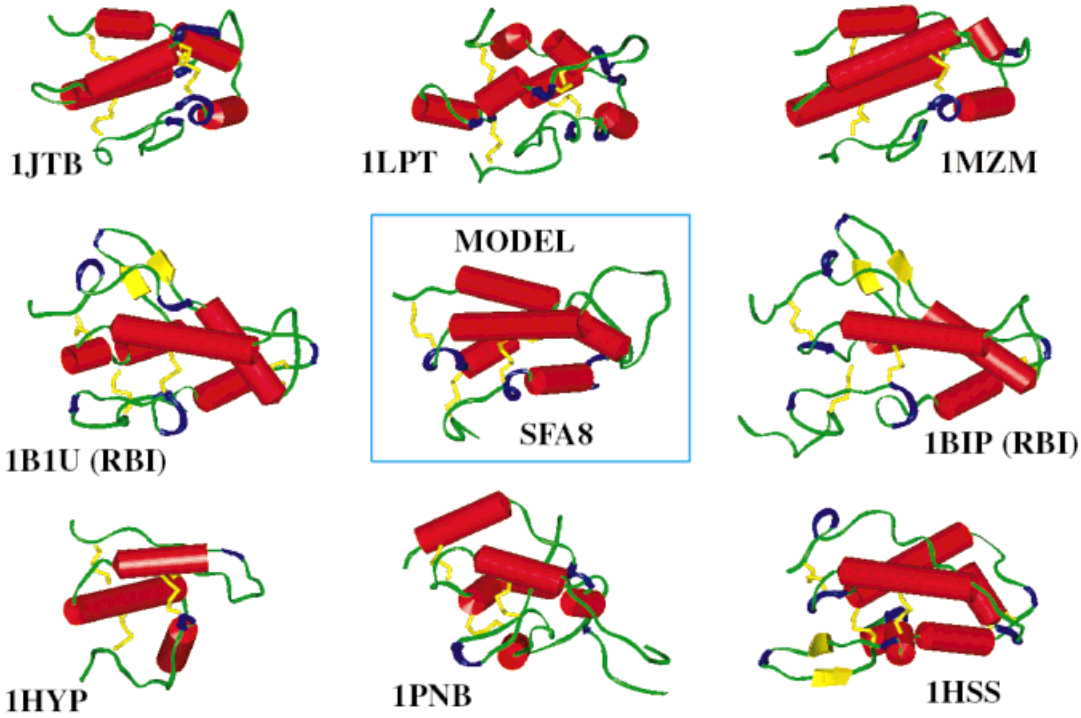


Fig. 1. Structures of the members of the SCOP "4 helices; folded leaf; right-handed superhelix; disulphide-rich" family and the model of SFA-8. Structures are labeled according to their Protein Database codes. Secondary structure is assigned according to Kabsch and Sander. Red cylinders are α-helices, yellow arrows β-strands, blue arrows are turns, green ribbons are loops, and yellow sticks are disulphide links. The view of each molecule corresponds to that obtained by superimposing the common cysteine residues for each protein on the SFA-8 model.

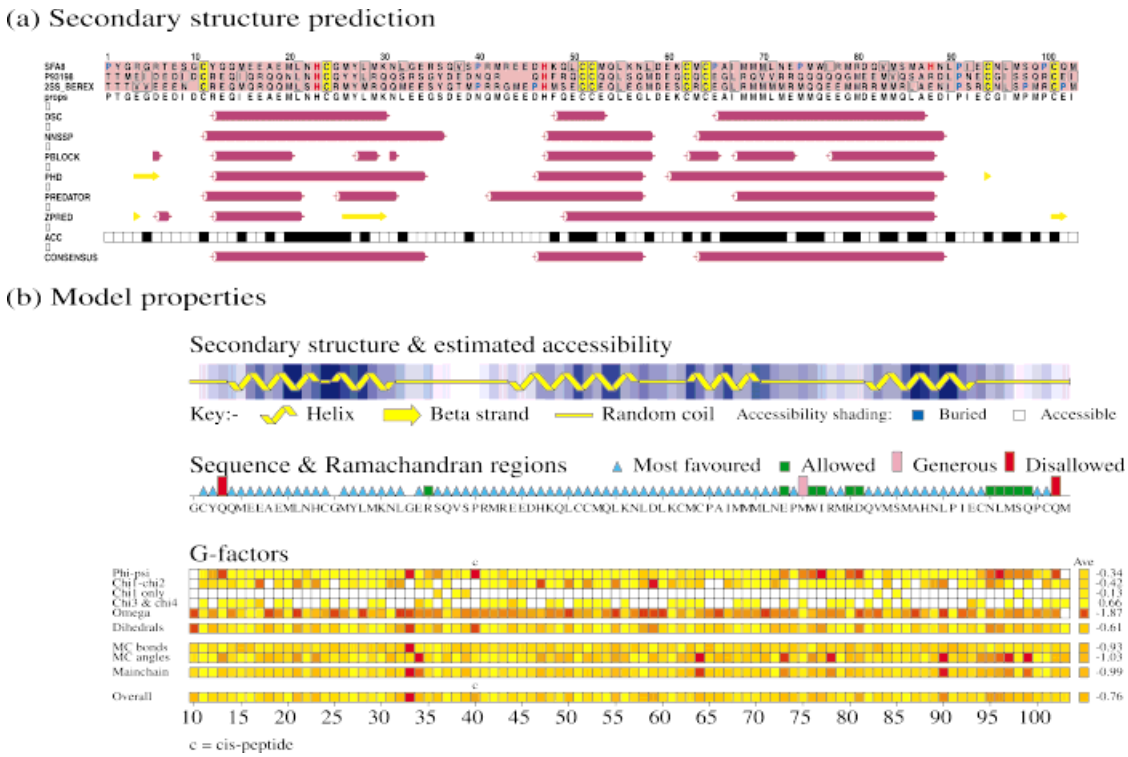


Fig. 2. **a:** A consensus secondary structure prediction of SFA-8 generated by the JPred server. **b:** A selection from the output generated by a Procheck analysis of the model SFA-8 structure.

Fig. 3. **a:** Clustal_W alignment of the amino acid sequences of SFA-8[5] and the seven structures of its assigned SCOP fold family. **b:** This alignment readjusted according to inspection of the seven structures.

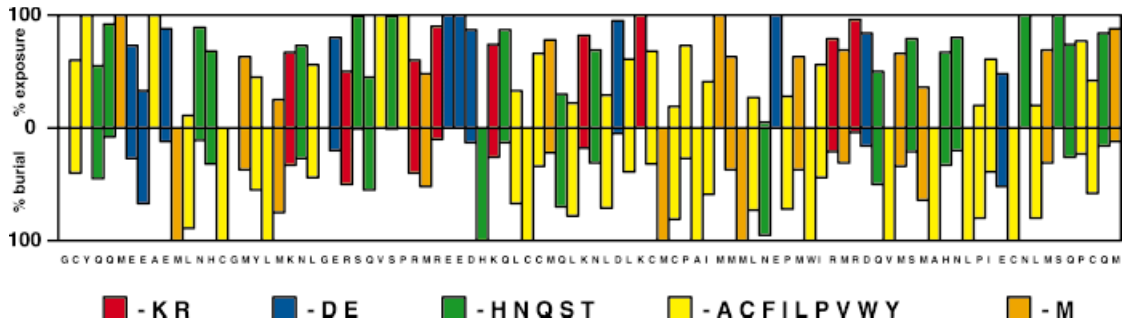


Fig. 4. Percentage surface exposures and burials of all residue side-chains in the model of SFA-8. Surfaces are calculated according to Connolly,[38] and percentages were calculated from surfaces determined for a particular residue in an extended polyglycine chain.

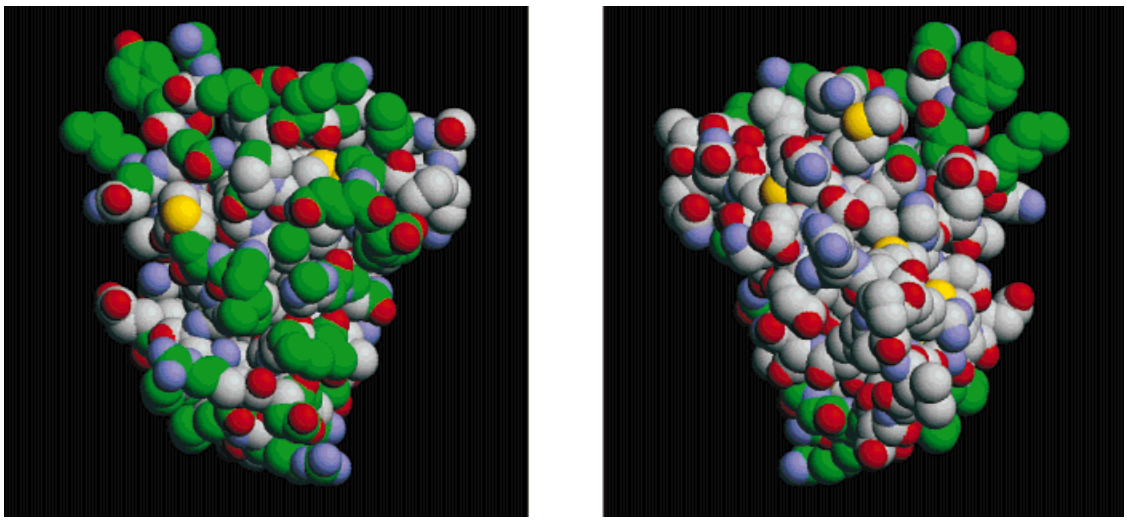

Fig. 5. Space-filling model of SFA-8 depicts surface hydrophobic patches. Rotation of the structure by 180° shows that SFA-8 has both an apolar (left) and a polar face (right). Atoms are colored: carbon and sulphur atoms involved in hydrophobic patches (green), other carbon (white), nitrogen (blue), oxygen (red), and other sulphur (yellow).

and a relatively large proportion of hydrophobic residues (16 Met, 9 Leu, 3 Ile, 3 Tyr, 2 Val, 1 Trp)[5].

Comparison of the amino acid sequence of SFA-8 with sequences in the SCOP database showed homology with a recently discovered class of protein whose structures contain four α-helices forming a right-handed superhelix.[25] The proteins constituting this group, and their pdb codes, are: RBI, a bifunctional α-amylase/trypsin inhibitor from ragi (Indian finger millet) seed[9] (1BIP and 1B1U); an α-amylase inhibitor from wheat kernel[26] (1HSS); a 2 S storage protein from rapeseed[8] (1PNB); a hydrophobic protein from soybean[25] (1HYP); and nonspecific lipid-transfer proteins (LTPs) from wheat seed[27] (1LPT), maize seed[28,29] (1MZM), barley seed[30] (1JTB), rice seed[31] (1RZL), and onion seed[32] (no structure deposited). The latter two structures were published since the current version of the SCOP database; both are structurally similar to the maize lipid transfer protein, 1MZM. The RBI, consisting of 122 amino acids with five disulphide bonds, shows the highest sequence identity to SFA-8 (24%), whereas the LTPs are less closely related (Fig. 3 and Table I).

Despite the low sequence identity amongst this set of proteins, they all contain a highly conserved skeleton of cysteine residues which is probably a major determinant of

their similar structures. These α-class proteins contain at least eight cysteine residues spaced in the following pattern:

$$X_{2-10}CX_{5-13}CX_{13-26}CCX_{9-19}CXCX_{22-45}CX_{6-13}CX_{2-8}$$

where X represents intervening amino acids and the subscript denotes the number of residues present in the different proteins. The disulphide bond pattern for SFA-8 has been determined by chemical methods as 11-62, 24-(51/52), (51/52)-94, and 64-101, with the connectivities of cysteine residues 51 and 52 undetermined.[6] NMR constraints map the disulphide pattern for RBI as 6-55, 20-44, 45-103, 57-114, and an additional bond 29-85. This is in perfect agreement with the disulphide pattern assigned unambiguously for a related α-amylase inhibitor from wheat kernel by mass spectrometry.[33] The NMR structures of the LTPs from wheat, maize, and barley, together with a high-resolution crystal structure of the maize LTP, show the disulphide bonds involving the fifth and sixth cysteine residues (which are separated by a single residue as in SFA-8) to be reversed, relative to the arrangement in RBI.

A model of the three-dimensional structure of SFA-8 was made by building the amino acid sequence of SFA-8 into the structure of RBI as described in Materials and Methods. The additional disulphide bond (29-85) in RBI is involved in stabilizing the trypsin inhibitory loop (Fig. 8a). The model of SFA-8 (Fig. 8b) has 50% of residues in α-helical conformation and lacks any β-structure (assuming that the first nine residues, which were not modeled, have no secondary structure), which is consistent with the secondary structure analyses by CD (~40%) and NMR spectroscopy (~50%). The model structure is composed of a bundle of four α-helices connected by short flexible loops and tightly linked by the four disulphide bonds. The apparent break in the N-terminal helix visible in Figures 1 and 2b is a result of the phi/psi values for Cys 24 falling just outside the Kabsch and Sander criteria for a helix residue, and is a distortion rather than a break of this helix. The α-helices are arranged in antiparallel pairs, and the four disulphide bonds are located in pairs on different sides of the structure. The end of the third helix is disulphide-linked to the C-terminus and to the N-terminal region, and the second helix is disulphide-linked to the first helix and to the end of the fourth helix.

Inspection of the model of SFA-8 shows that the protein has an orthodox hydrophobic core, and every charged group on side chains of Arg, Lys, Glu, and Asp are fully solvated. This is expressed in a semiquantitative manner in Figure 4, which shows the percentage exposure of all residue side chains in the model. Analysis of the protein surface for the presence of hydrophobic patches reveals that the molecule possesses one polar and one apolar face. The program, Quilt[13], delineates continuous areas of solvent-accessible surface formed exclusively from neighboring nonpolar atoms (i.e., either carbon or sulphur).[34] The six largest hydrophobic patches identified using Quilt, with solvent-accessible surface areas ranging from 176 to 373 Å$^2$, are depicted in Figure 5. Several residues, mostly from the two C-terminal α-helices and the connecting loop of the model structure (including the hydrophobic residues Met 15, Leu 32, Leu 58, Ile 67, Met 68, Met 69, Leu 71, Met 75, Ile 77, Met 79, Met 84, Met 86, Ile 92, Leu 96, Met 97, and Met 103) combine to form this unusual apolar face containing no charged residues.

### Structural Information From Experiment

Native SFA-8 was purified from sunflower seeds by cation exchange chromatography and RP-HPLC. The molecular weight of the purified SFA-8 is 12,148 (±3) as determined by electrospray mass spectrometry. This value agrees with the value of 12,147.48 calculated from the published amino acid sequence[5] with four intact disulphide bonds and confirms the identity and integrity of the purified protein.

The association state of SFA-8 was determined by analytical ultracentrifugation. Protein solutions at 20, 50, and 100 μmol/L in pH 7 buffer were clearly monomeric with molecular weights of 12,350 (±250) g mol$^{-1}$, 11,820 (±490) g mol$^{-1}$, and 11,400 (±125) g mol$^{-1}$, respectively. However, at a concentration of 2 mmol/L, the z-average molecular weight of 17,100 (±100) g mol$^{-1}$ obtained from sedimentation equilibrium measurements is greater than the monomer molecular weight of 12,147 g mol$^{-1}$, implying that the protein is capable of a degree of self-association at these elevated concentrations. The weight obtained from sedimentation equilibrium of 2.5 mmol/L SFA-8 at pH 4.5 is 20,500 (±370) g mol$^{-1}$, indicating that the protein undergoes reversible association.

Far-UV (amide) CD spectroscopy of the native protein at 25°C (Fig. 9a) showed a maximum at 190 nm and minima at 209 nm and 221 nm, which are characteristic of a protein with a high content of α-helical structure (about 40% from the signal at 222 nm[17]).

Two-dimensional NMR spectra of native SFA-8 show good spectral dispersion and resolution at 27, 40, and 50°C, and especially 70°C. The NH-NH region of a NOESY spectrum at 70°C (Fig. 9b) shows a large number of cross-peaks. Almost all of the CHα signals identified in COSY/HOHAHA spectra have chemical shifts below 4.8 ppm (data not shown). According to the Chemical Shift Index of Wishart and Sykes,[35] the identifiable CHα are likely to be from protons in a helical structure. Therefore, it is assumed that the NH-NH NOESY cross-peaks are largely sequential $NH_i$-$NH_{i+1}$ cross peaks in helix. At least 50 such NH-NH cross-peaks are observed, indicating a helical content of at least 50%.

Efforts to extend the NMR analysis to three-dimensional structure determination were hindered by the loss of off-diagonal signals in DQF-COSY experiments at temperatures of 50°C and below. The fingerprint region of DQF-COSY spectra (data not shown), which should contain ~100 signals representing chemical coupling between the backbone amide proton and alpha carbon proton(s) of each residue in the SFA-8 molecule, shows considerable loss of signals (only ~60 signals could be identified). Although a degree of spectral overlap and the loss of

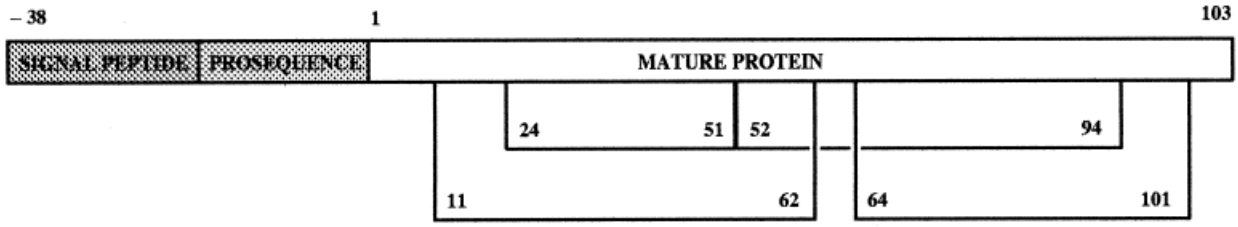


Fig. 6. Disulphide pattern of SFA-8 determined by chemical methods.[6]

cross-peaks at the chemical shift of the water resonance could account for some loss of signals, this number is much lower than expected. The fingerprint regions of both the DQF-COSY and HOHAHA spectra clearly show strong peaks representing a subset of residues within SFA-8.

Loss of COSY cross-peaks may indicate that the coupling constant is smaller than the intrinsic linewidth. These effects are normally attributed to solvent viscosity or aggregation caused by high sample concentration, which reduces the rate of tumbling of the protein in solution. As stated earlier, sedimentation analysis at pH 4.5 and 20°C demonstrates that SFA-8 self-associates at concentrations used for these NMR measurements (≥2 mmol/L). Consistent with this finding, markedly improved spectral data were obtained from two-dimensional homonuclear experiments at 70°C. However, additives expected to suppress self-association (20% TFE, 2 mol/L GdmCl, 1.5 mol/L GdmSCN, 20 mmol/L CHAPS detergent) failed to improve the spectra. This suggests that either the loss of signal results from the interconversion of a heterogeneous population of conformations on the time scale of the NMR chemical shift or coupling constant (1–500 $s^{-1}$) or that these additives were unable to remove self-association

The presence of hydrophobic surface patches is consistent with the strong retention of SFA-8 on a reversed-phase HPLC column[24] and is probably responsible for the highly stable emulsions formed by SFA-8 at concentrations of 0.5–1 mg $ml^{-1}$ in oil/water mixtures.[2] Emulsions are thermodynamically unstable mixtures of immiscible liquids, and most soluble proteins will decrease the surface tension and increase resistance to coalescence by adsorption to emulsion droplets.[36] Proteins containing a large number of hydrophobic residues are good emulsifiers, and the accessibility of the hydrophobic groups at the surface of the adsorbed protein is important.[3] Therefore, residues located in hydrophobic patches on protein surfaces are likely to participate in emulsification.[4] Reduced SFA-8 is expected to display a larger exposed hydrophobic surface because of destabilization of the protein fold, explaining why the emulsifying properties are improved by the presence of reducing agent.[37]

Protein digestion can provide information on the conformation of the substrate protein and the location of susceptible peptide bonds. Native SFA-8 underwent cleavage at a single site, determined as Arg 41-Met 42 by N-terminal sequencing data. This site is located on a surface loop in

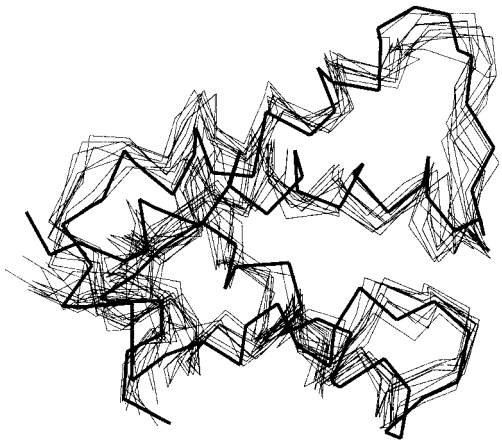

Fig. 7. A superimposition of Cα traces of the 10 structures obtained by simulated annealing on the initial minimized model of SFA-8 (bold line).

the model of SFA-8, as indicated in Figure 8b. The C-terminal fragment of cleaved SFA-8 appears as a lower molecular weight band (relative to intact SFA-8) during SDS-PAGE that increases in intensity with time until all of the protein sample is cleaved (10 min) and still persists after 60 min incubation (data not shown). An additional band corresponding to the N-terminal fragment disappears with time because of further digestion.

Despite the low sequence homology of SFA-8 with other members of the "4 helices; folded leaf; right-handed superhelix; disulphide-rich" fold class, the distinctive disulphide pattern of the protein places it firmly in this family. An homology model was built using RBI as the template structure, because this protein has the highest sequence identity with SFA-8. The α-helical content of the model (50%) compares favorably with that derived from CD (40%) and NMR (50%) experiments. The model has acceptable geometry, a hydrophobic core, and solvated charged residues. It exhibits one unusually hydrophobic face, consistent with the emulsifying properties of the protein, and the trypsin cleavage site present in SFA-8 is on a solvent-exposed surface loop. Hence, the model is consistent with all the available experimental data.

## CONCLUSION

In summary, sequence homology and disulphide bond pattern show that SFA-8 is related to a class of protein whose structures comprise an antiparallel bundle of four

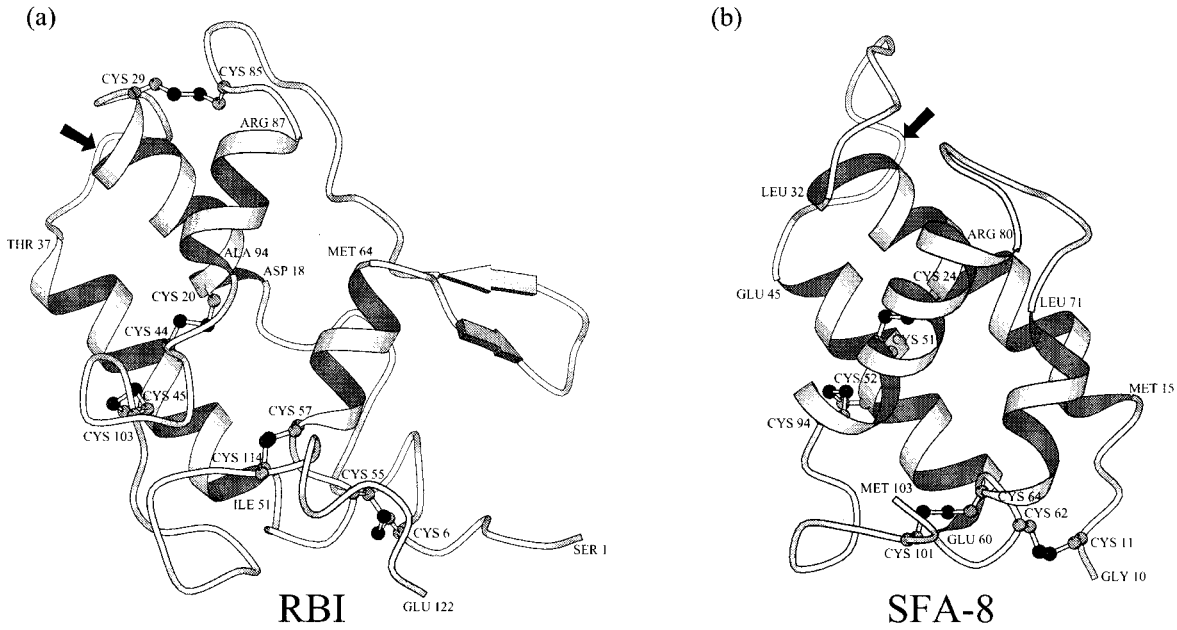


Fig. 8. A comparison of the template structure for homology modeling, RBI **(a)** and the homology model of SFA-8 **(b),** generated using Molscript.[39] The trypsin inhibitory site in RBI and the trypsin cleavage site in SFA-8 are indicated with an arrow.

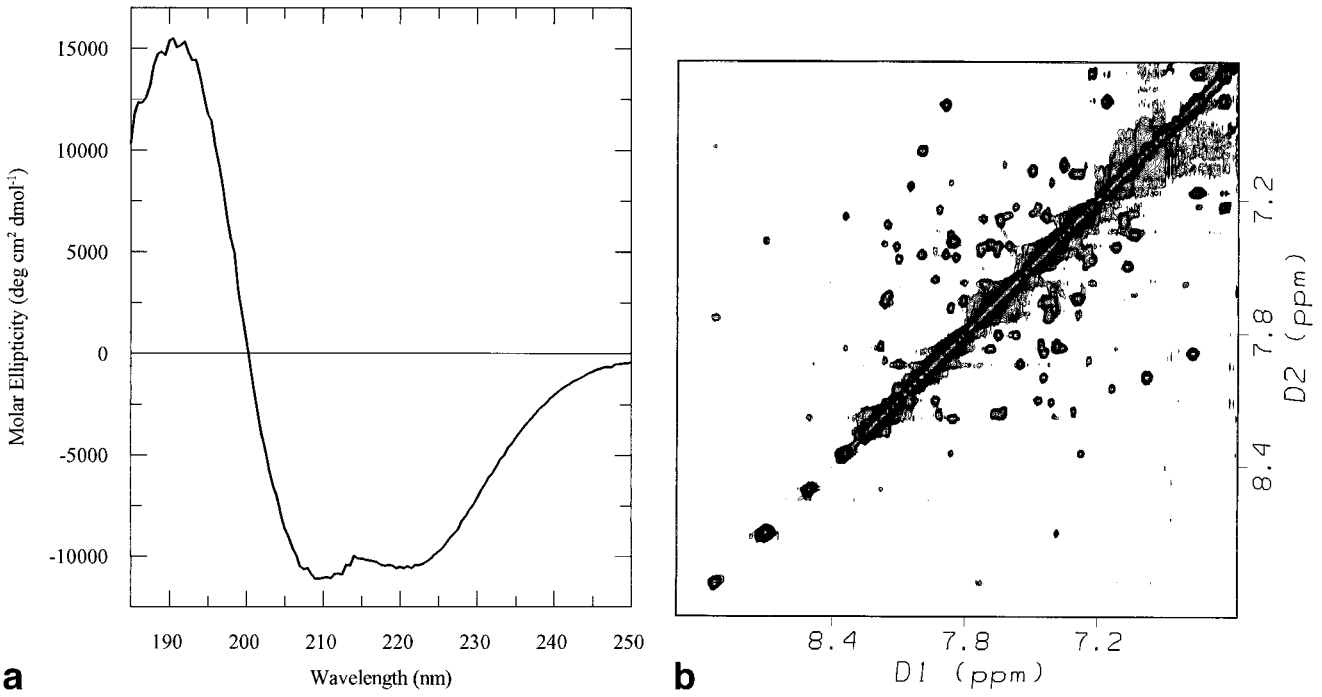


Fig. 9. Secondary structure determination: **a:** Far-UV (amide region) circular dichroism spectroscopy of SFA-8 at 25°C. **b:** NMR spectroscopy of SFA-8 showing the NH-NH region of a 500-MHz NOESY spectrum at pH* 4.5 and 70°C.

or five α-helices cross-linked by disulphide bonds. Commensurate with this finding, CD and NMR spectroscopy demonstrate that SFA-8 has a high content of α-helical structure. Homology modeling of the protein structure suggests that SFA-8 possesses a large hydrophobic face, consistent with its good emulsifying properties.

## ACKNOWLEDGMENTS

The authors thank the following for assistance with these experiments: R. J. Fido and N. A. Ranson (purification of SFA-8), J. Takei (NMR experiments), N. Errington at the National Centre for Macromolecular Hydrodynam-

ics at Leicester University (analytical ultracentrifugation), G. Bloomberg (mass spectrometry), and Z. Patterson (N-terminal sequencing). We thank the Biotechnology and Biological Sciences Research Council for CASE project funding and grant-aided support to IACR. A.R.C. is a Lister Institute Fellow. We thank M. J. Parker, P. M. Wood, and B. Beaumont for critical assessment of the manuscript.

## REFERENCES

1. Mazhar H, Quayle R, Fido RJ, Stobart AK, Napier JA, Shewry PR. Synthesis of storage reserves in developing seeds of sunflower. Phytochemistry 1998;48:429–432.
2. Guéguen J, Popineau Y, Anisimova IN, Fido RJ, Shewry PR, Tatham AS. Functionality of the 2 S albumin seed storage proteins from sunflower (*Helianthus annuus* L.). J Agric Food Chem 1996;44:1184–1189.
3. Fligner KL, Mangino ME. Relationship of composition to protein functionality. ACS Symposium Series 1991;454:1–12.
4. Li-Chan E, Nakai S. Importance of hydrophobicity of proteins in food emulsions. ACS Symp Series 1991;448:193–212.
5. Kortt AA, Caldwell JB, Lilley GG, Higgins TJV. Amino acid and cDNA sequences of a methionine-rich 2 S protein from sunflower seed (*Helianthus annuus* L.). Eur J Biochem 1991;195:329–334.
6. Egorov TA, Odintsova TI, Musolyamov AK, Fido R, Tatham AS, Shewry PR. Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily. FEBS Lett 1996;396:285–288.
7. Shewry PR. Plant storage proteins. Biol Rev 1995;70:375–426.
8. Rico M, Bruix M, González C, Monsalve RI, Rodríguez R. $^1$H NMR assignment and global fold of napin BnIb, a representative 2 S albumin seed protein. Biochemistry 1996;35:15672–15682.
9. Strobl S, Mühlhahn P, Bernstein R, et al. Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin inhibitor from ragi seeds by NMR spectroscopy. Biochemistry 1995;34:8281–8293.
10. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
11. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998;14:892–893.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
13. Lijnzaad P, Berendsen HJC, Argos P. A method for detecting hydrophobic patches on protein surfaces. Proteins 1996;26:192–203.
14. Kortt AA, Caldwell JB. Low molecular weight albumins from sunflower seed: identification of a methionine-rich albumin. Phytochemistry 1990;29:2805–2810.
15. Schägger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 1987;166:368–379.
16. Lamm O. Die Differentialgleichung der Ultrazentrifugierung. Ark Mat Astron Fys 1929;21B:2:1–4.
17. Chen Y, Martinez HM. Determination of the secondary structure of proteins by circular dichroism and optical rotatory dispersion. Biochemistry 1972;11:4120–4131.
18. Morris GA, Freeman R. Selective excitation in Fourier transform nuclear magnetic resonance. J Magn Reson 1978;29:433–462.
19. Wider G, Macura S, Kumar A, Ernst RR, Wüthrich K. Homonuclear 2-dimensional $^1$H NMR of proteins—experimental procedures. J Magn Reson 1984;56:207–234.
20. Rance M. Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. J Magn Reson 1987;74:557–564.
21. Rance M, Sorenson OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. Improved spectral resolution in COSY $^1$H NMR spectra of proteins via double quantum filtering. Biochem Biophys Commun 1983;117:479–485.
22. Klevit RE. Improving two-dimensional NMR spectra by T1 ridge subtraction. J Magn Reson 1985;62:551–555.
23. Anglister J, Grzesiek S, Ren H, Klee CB, Bax A. Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. J Biomol NMR 1993;3:121–126.
24. Anisimova IN, Fido RJ, Tatham AS, Shewry PR. Genotypic variation and polymorphism of 2 S albumins of sunflower. Euphytica 1995;83:15–23.
25. Baud F, Pebay-Peyroula E, Cohen-Addad C, Odani S, Lehmann MS. Crystal structure of a hydrophobic protein from soybean; a member of a new cysteine-rich family. J Mol Biol 1993;231:877–887.
26. Oda Y, Matsuraga T, Fukuyama K, Miyazaki T, Morimoto T. Tertiary and quaternary structures of 0.19 α-amylase inhibitor from wheat kernel determined by x-ray analysis at 2.06 Å resolution. Biochemistry 1997;36:13503–13511.
27. Gincel E, Simorre J-P, Caille A, Marion D, Ptak M, Vovelle F. Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional 1H-NMR data. Eur J Biochem 1994;226:413–422.
28. Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. Structure 1995;3:189–199.
29. Gomar J, Petit M-C, Sodano P, et al. Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. Protein Sci 1996;5:565–577.
30. Heinemann B, Andersen KV, Nielsen PR, Bech LM, Poulsen FM. Structure in solution of a four-helix lipid binding protein. Protein Sci 1996;5:13–23.
31. Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW. Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. J Mol Biol 1998;276:437–448.
32. Tassin S, Broekaert WF, Marion D, et al. Solution structure of Ace-AMP1, a potent antimicrobial protein extracted from onion seeds. Structural analogies with non-specific lipid transfer proteins. Biochemistry 1998;37:3623–3637.
33. Poerio E, Caporale C, Carvano L, Pucci P, Buonocore V. Assignment of the five disulfide bridges in an α-amylase inhibitor from wheat kernel by fast-atom-bombardment mass spectrometry and Edman degradation. Eur J Biochem 1991;199:595–600.
34. Eisenhaber F, Argos P. Hydrophobic regions on protein surfaces: definition based on hydration shell structure and a quick method for their computation. Protein Eng 1996;9:1121–1133.
35. Wishart DS, Sykes BD. Chemical-shifts as a tool for structure determination. Mets Enzymol 1994;239:363–392.
36. Halling PJ. Protein-stabilized foams and emulsions. CRC Crit Rev Food Sci Nutr 1981;15:155–203.
37. Popineau Y, Tatham AS, Shewry PR, Marion D, Guéguen J. 2 S sunflower albumins: functional properties of native and modified proteins. In: Guéguen J, Popineau Y, editors. Plant proteins from European crops: food and non-food applications. Berlin: Springer-Verlag; 1998. p 131–135.
38. Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983;16:548–558.
39. Kraulis PJ. Molscript: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–960.