Solution Structure of a Methionine-Rich 2S Albumin from Sunflower Seeds: Relationship to Its Allergenic and Emulsifying Properties[†,‡]

David Pantoja-Uceda,[§] Peter R. Shewry,[||] Marta Bruix,[§] Arthur S. Tatham,[||] Jorge Santoro,[§] and Manuel Rico*[,§]

Instituto de Química Física "Rocasolano", CSIC, Serrano 119, Madrid 28006, Spain, and Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, U.K.

Received February 11, 2004; Revised Manuscript Received April 6, 2004

ABSTRACT: The three-dimensional structure in aqueous solution of SFA-8, a 2S albumin 103-residue protein from seeds of sunflower (*Helianthus anuus* L.), has been determined by NMR methods. An almost complete ^{1}H resonance assignment was accomplished from analysis of two-dimensional (2D) COSY and 2D TOCSY spectra, and the structure was computed by using restrained molecular dynamics on the basis of 1393 upper limit distance constraints derived from NOE cross-correlation intensities measured in 2D NOESY spectra. In contrast with most other 2S albumins, SFA-8 consists of a single polypeptide chain without any cleavage in the segment of residues 30−46. The computed structures exhibited an rmsd radius of 0.52 Å for the backbone structural core (residues 11−30 and 46−101) and 1.01 Å for the side chain heavy atoms. The resulting structure consists of five amphipathic helices arranged in a right-handed superhelix, a folding motif first observed in nonspecific lipid transfer (nsLTP) proteins, and common to other 2S albumins. In contrast to nsLTP proteins, neither SFA-8 nor RicC3 (a 2S albumin from castor bean) has an internal cavity that is able to host a lipid molecule, which results from an exchange in the pairing of disulfide bridges in the CXC segment. Both 2S albumins and nonspecific lipid transfer proteins belong to the prolamin superfamily, which includes a number of important food allergens. Differences in the extension and solvent exposition of the so-called "hypervariable loop" (which connects helices III and IV) in SFA-8 and RicC3 may be responsible for the different allergenic properties of the two proteins. SFA-8 has been shown to form highly stable emulsions with oil/water mixtures. We propose that these properties may be determined partly by a hydrophobic patch at the surface of the protein which consists of five methionines that partially hide the Trp76 residue. The flexibility of the loop which contains Trp76 and the hydrophobicity of the whole environment may favor a conformational change, by which the Trp76 side chain may become inserted into the oil phase.

2S albumins are cysteine-rich water-soluble proteins present in the seeds of a wide range of dicotyledonous plants with sedimentation coefficients of approximately 2S (*1*, *2*). They are small proteins (12−15 kDa), with a characteristic distribution of eight cysteins in a conserved pattern (...C...C.../...CC...CXC...C...C...), and are generally composed of two different polypeptide chains, of ∼3−5 and ∼8−10 kDa, linked by two disulfide bridges. They also have two additional intrachain bonds, which make the proteins very stable and compact. 2S albumins form a family of storage proteins belonging to the prolamin superfamily (*3*, *4*) that includes cereal seed inhibitors of α-amylase and/or trypsin, puroindolines, nonspecific lipid transfer proteins (nsLTPs),[1] the hydrophobic protein from soybean, and some cell wall glycoproteins. Most 2S albumins appear to function solely as storage proteins, which is consistent with their amino acid compositions and their patterns of accumulation during seed development and degradation during germination. However, other activities have been ascribed to some 2S albumins, which have been shown to act as antifungal agents (*5*−*7*) and inhibitors of serine proteases (*8*−*10*). This is consistent with a secondary role of these proteins in defense. In addition, the 2S albumins have been used for genetic engineering, as carriers for the synthesis of biologically active peptides (*11*), and to improve the nutritional properties of grain crops by increasing their contents of essential amino acids (*12*). The prolamin superfamily of proteins has also been the subject of recent interest because of the fact that a number of its members, including many 2S albumins, are allergens (*13*−*16*). This is of particular interest in relation to the potential allergenicity of novel and transgenic foods (*17*).

We have been studying the three-dimensional structures of different members of the 2S albumin family to obtain information about the structural basis for their antimicrobial

[†] This work was supported by Project PB98-0667 of the DGESIC of the Spanish Ministry of Education and Project BIO2002-00720 of the Spanish Ministry of Science and Technology granted to M.R. and J.S., respectively. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. D.P.-U. was supported by a predoctoral grant from the Spanish Ministry of Science and Technology.

[‡] Coordinates for the minimized 20 structures of SFA-8 have been deposited in the RCSB Protein Data Bank (entry 1S6D).

* To whom correspondence should be addressed. Fax: +34-91-5642431. Telephone: +34-91-5649400. E-mail: mrico@iqfr.csic.es.

[§] CSIC.

[||] Rothamsted Research.

[1] Abbreviations: nsLTP, nonspecific lipid transfer protein; rmsd, root-mean-square deviation; SFA-8, sunflower albumin 8.

10.1021/bi0496900 CCC: $27.50
Published on Web 05/11/2004

activity and allergenicity. A global fold was first determined for the napin Bn1b (*18*) isolated from rapeseed, and then for recombinant ^{15}N-labeled RicC3 from *Ricinus communis*, for which a well-defined three-dimensional structure could be determined (*19*). The resulting global structure, which was defined as a four-helix bundle arranged in a right-handed superhelix (*20*), is shared by other members of the prolamin superfamily. However, some small differences are observed between the structures of the different protein families, which may relate to their different properties.

The 2S albumins of sunflower (*Helianthus annuus* L.) account for approximately one-third of the total salt-soluble proteins in the seed. The major methionine-rich sunflower albumin, called SFA-8, has been particularly interesting to one of the contributing laboratories, because of its ability to form highly stable emulsions with oil/water mixtures (*21*). SFA-8 contains an unusually high proportion of hydrophobic residues, including 16 methionines, in a mature protein of 103 residues. In contrast with other 2S albumins, SFA-8 and other sunflower albumins are not cleaved *in planta* into large and small subunits but remain as single polypeptide chains. SFA-8 exhibits no tryptic, chymotryptic, or α-amylase inhibitory activity *in vitro*, despite the presence of a putative trypsin inhibitory site (*22*). Recently, it has been demonstrated that SFA-8 may also be allergenic (*23*, *24*), although this result is in conflict with the low prevalence of sunflower seed allergy in countries such as Spain which have high levels of consumption of raw seeds. The ability to prepare highly purified SFA-8, free from any residue heterogeneity, prompted us to determine its three-dimensional structure in aqueous solution by NMR methods and to elucidate the molecular basis for its emulsification properties and putative allergenicity.

MATERIALS AND METHODS

Protein Sample. SFA-8 was purified from sunflower seeds as described in refs *25* and *26*. The amino acid sequence was confirmed by MALDI-tof and Q-tof MS of the whole protein and specific peptides.

NMR Spectroscopy. NMR experiments were performed on samples of ~1.0 mM protein dissolved in 95% H_2O and 5% D_2O or in a D_2O solution containing 10 mM sodium phosphate and TSP, at pH 3.0 in 5 mm NMR tubes. All NMR spectra were recorded at 308 K on Bruker AV 600 and Bruker AV 800 NMR spectrometers operating at 600.13 and 800.10 MHz for ^{1}H observation, respectively. Proton chemical shifts were referenced to internal trimethylsilyl propionate (TSP). Two-dimensional (2D) COSY, TOCSY (mixing time of 60 ms), and NOESY (mixing times of 60 and 80 ms) spectra were obtained for resonance and NOE assignments. Water suppression was achieved either by selective presaturation or by including a WATERGATE module (*27*) in the original pulse sequences prior to acquisition. Conventional pulse sequences and phase cycling procedures were used in all experiments (*28*). To test the association state of the sample, one-dimensional (1D) ^{1}H spectra were run at sample concentrations of 1 and 0.1 mM at pH 3.0 and 308 K. No changes in chemical shifts or line widths were observed upon 10-fold dilution, so we concluded that the sample was monomeric under the conditions described above.

The size of the acquisition data matrix for the 2D spectra was 2048 × 512 or 4096 × 512 data points in the direct or indirect dimension, respectively, when operating at 600 or 800 MHz. The data matrices were zero filled to duplicate the number of data points in all dimensions. Resolution enhancement methods included sine-bell or square-sine-bell windows with optimized shifts for every spectrum. Baseline correction was applied in both dimensions.

The processing of the 2D spectra was carried out by using XWINNMR (Bruker Biospin, Karlsruhe, Germany), and the spectral analysis, assignment, and cross-peak volume calculations were performed with XEASY (*29*).

Proton resonances were assigned by combining information extracted from 2D experiments, using the standard strategy described in ref *30*. Data have been deposited in the BioMagResBank database (*31*) under accession number BMRB-6082.

Distance and Angle Constraints. After the assignment of most proton resonances belonging to both the main chain and the side chains, the 2D NOESY spectra were analyzed to assign as many dipolar cross-correlations as possible. They were converted, after internal calibration, into upper distance limits with CALIBA (*32*). The limits were restricted within the range of 2.7−6.0 Å. The strongest sequential HN_i-HN_{i+1} connectivity in a helical region was taken to be the reference for the 2.8 Å distance. Finally, after DYANA filtering, 1393 upper limit distance constraints were derived, distributed in 396 intraresidual, 406 sequential, 335 short-range, and 256 long-range constraints. Additionally, as the SFA-8 contains eight cysteine residues, four sets of standard upper and lower limits for each identified disulfide bond [2.1/2.0 Å for $S\gamma(i)-S\gamma(j)$ and 3.1/3.0 Å for $C\beta(i)-S\gamma(j)$ and $C\beta(j)-S\gamma(i)$] were introduced during the rounds of calculations. The distance constraints were supplemented with 14 χ_1 torsion angle constraints derived from a qualitative analysis of the COSY $H\alpha-H\beta$ cross-peaks. No stereospecific assignments of prochiral protons were introduced in the initial calculations, and the usual correction for pseudoatoms (*30*) was added in all cases. In the final steps, the stereospecific assignments of $H\beta-H\beta'$ resonances of residues Tyr12, Tyr27, Leu28, Arg43, Leu50, Cys62, Met69, Asn72, Trp76, Met86, Asn89, Leu90, Cys94, and Cys101, the $H\delta$ resonance of Pro65, the $Q\delta$ resonance of Leu90, and the $Q\gamma$ resonance of Val83 were generated with GLOMSA (*32*) on the basis of the set of 20 protein structures and interproton distance upper limits.

Structure Calculation. The three-dimensional solution structure was determined by simulated annealing combined with molecular dynamics in torsion angle space followed by a final minimization.

Recently, a high-resolution three-dimensional structure of rRicC3 has been reported (*19*), which constitutes a suitable representative structure for the broad family of seed 2S albumin proteins. Although rRicC3 and SFA-8 have a low level of sequence identity (~20%), we built a three-dimensional (3D) model for SFA-8 with MODELLER 6 version 1 (*33*) by using the 3D structure of rRicC3 as a template. The 3D model was used to make a rough screening of tentative NOE assignments based only on chemical shifts. The assignments involving protons separated by a much larger distance than the upper limit distance were eliminated in the first round of the NOE assignment automated procedure. In this way, the time-consuming process of NOE identification in the NOESY spectra is greatly alleviated. To

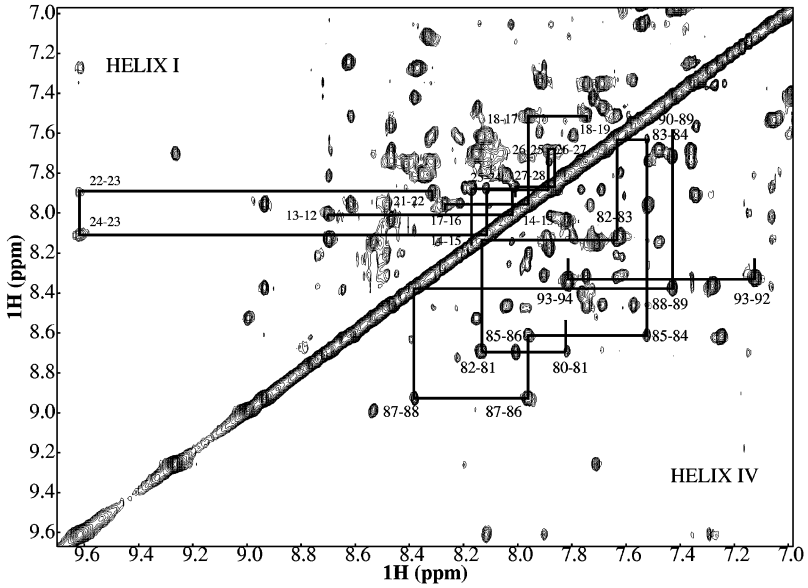

FIGURE 1: Amide−amide region of the 2D NOESY 800 MHz spectrum, with links between sequential H^N-H^N cross-correlations of helices Ia and Ib above the diagonal and helix IV below the diagonal.

ensure that the structure calculation is not biased toward an incorrect structure, a small set of isolated and unambiguous NOE peaks were manually assigned from NOESY spectra. An initial structure calculation, followed by automatic peak NOE assignment using the 3D model and the unambiguous distance constraints, was performed with DYANA version 1.5 (*34*) and the algorithms for automatic assignment of NOESY spectra, NOAH (*35*) and CANDID (*36*). On the basis of the degree of consistency of these initial structures, an additional set of NOEs could then be assigned and subsequently used to calculate a new set of structures. Several such cycles of structure calculation and NOE assignment were used to assign the maximum number of NOE peaks that were possible. From these distance and angular constraints, a set of 50 structures was calculated using DYANA version 1.5. All His, Lys, and Arg side chains were treated as positively charged and all Glu and Asp side chains as negatively charged. The 20 structures with the lowest DYANA target were refined by energy minimization *in vacuo* (200 steps) in the presence of NMR constraints using AMBER7 (*37*). All DYANA and AMBER calculations were performed on a cluster of eight Intel Pentium II computers with the Linux operating system. PROCHECK-NMR (*38*) was used to analyze the quality of the refined structures and MOLMOL (*39*) to prepare images of the molecules.

RESULTS

Assignment. Assignments of backbone proton resonances are complete, with the exception of the HN and Hα resonances of Glu44 and Met103. A sample of the sequence-specific ^{1}H assignment is shown in Figure 1. Further analysis of the TOCSY and NOE data permitted a nearly complete ^{1}H side chain assignment with the only exceptions being some distal side chain resonances in arginines and lysines and some side chain NH_2 resonances of asparagines and glutamines. The percentage of assigned side chain proton resonances was of 92.5%. All five Xxx−Pro bonds present in the protein are in the *trans* conformation, based on intense NOESY Xxx(Hα)−Pro(Hδ*) sequential cross-peaks (*30*).

Disulfide Bonds. Four disulfide bonds stabilize the three-dimensional structure. The individual covalent links between the cysteine residues could be determined by careful analysis of the NOE spectra since the distance between Hβ of one Cys and Hα of the second Cys involved in the same disulfide bond is within the range of detectable NOEs (*40*). Several inter-cysteine NOE connectivities allowed us to clearly identify three disulfide bonds: Cys11(Hβ2)−Cys62(Hα) (SS1), Cys24(Hα)−Cys51(Hβ2,Hβ3) (SS2), and Cys52-(Hα)−Cys94(Hβ2) and Cys52(Hβ2)−Cys94(Hβ3) (SS3). Consequently, the fourth disulfide bond (SS4) must be a Cys64−Cys101 bond. Long-range NOEs between residues in the vicinity of the cysteines (see Figure 3B) confirm the disulfide pattern, which also agrees with those determined by peptide mapping of SFA-8 (*41*, *42*) and analyses of other members of the 2S albumin family (*18*, *19*, *41*).

Secondary Structure. Figure 2 summarizes the Hα conformational shifts, sequential and short-range NOEs, and NH exchange data. The conformational shifts are calculated with the equation $\Delta\delta_{H\alpha} = \delta_{H\alpha}{}^{OBS} - \delta_{H\alpha}{}^{RC}$, where $\delta_{H\alpha}{}^{RC}$ values are the shifts corresponding to the random coil (*43*), incorporating corrections for the effects of proline and oxidized cysteine on the chemical shifts of peptide backbone nuclei (*44*). The presence of five helical secondary structure elements was initially indicated by a dense array of characteristic NOEs and by stretches of slowly exchanging backbone amide protons. All these data support the presence of five α-helices spanning approximately residues 11−17, 22−30, 49−57, 63−71, and 79−93. Rounds of three-dimensional structure calculations led to more precise limits for the helices.

Tertiary Structure. A total of 1393 interproton upper distance constraints, corresponding to nonoverlapping and unambiguously assigned cross-correlations, were used in the calculation, as described in Materials and Methods. Figure 3B shows a diagonal plot of NOEs identified in the NOESY spectra. The global folding of the polypeptide chain is uniquely defined due to the large number of NOEs in the core of the protein (Figure 3A).

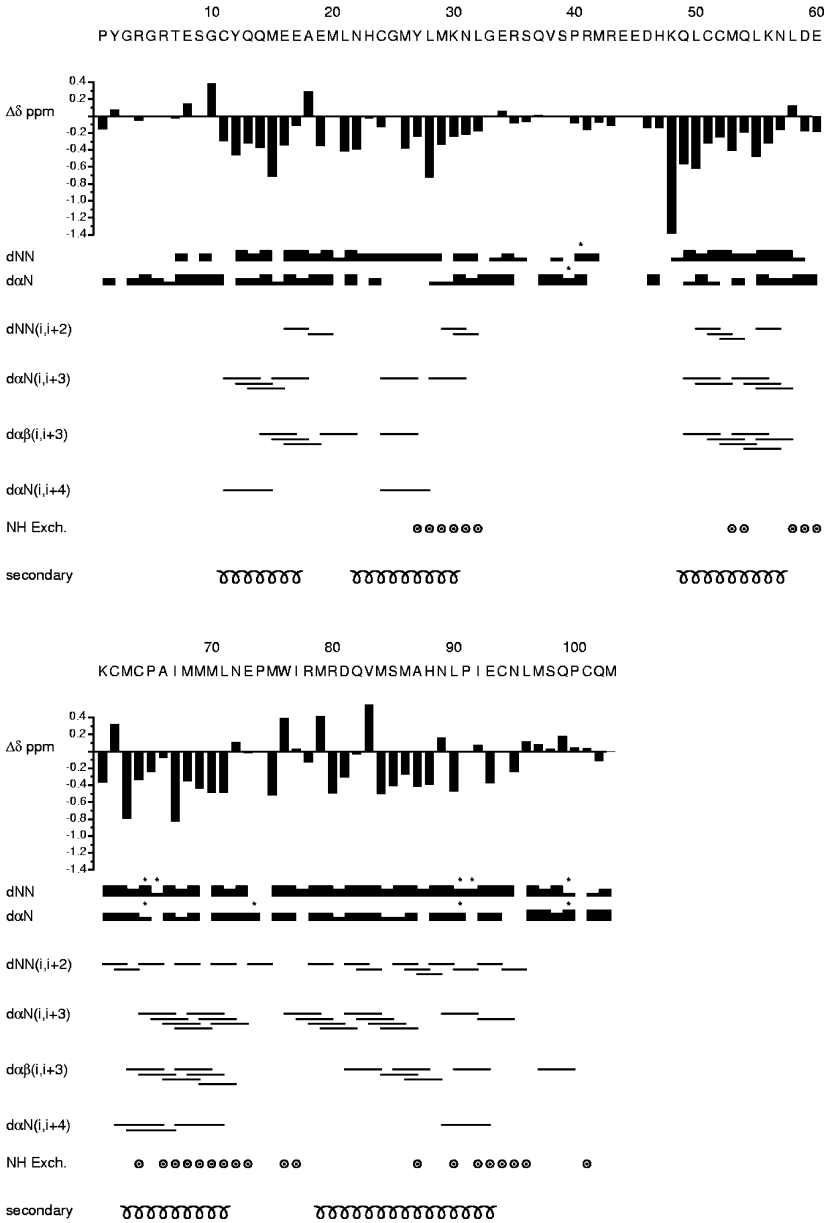

FIGURE 2: Summary of Hα conformational chemical shifts, short-range NOEs, and slowly exchanging amide protons. NOE sequential cross-correlations between backbone protons of residues adjacent to prolines and H$\delta\delta'$ proline protons (*trans* peptide bond conformation) are marked with an asterisk.

Statistical details about the quality and precision of the 20 best structures are summarized in Table 1, while selected structures are superimposed in Figure 4. The final 20 minimized structures are fully compatible with the resonance and NOE cross-correlation assignments, with no violations of distance constraints greater than 0.4 Å and a maximum violation of 0.35 Å. The average number of inter-residue distance constraints per residue was 13.5. The average rmsd from the mean for backbone atoms of the most structured regions are 0.52 and 1.01 Å for all heavy atoms, excluding unstructured regions at the N- and C-termini (residues 1−10 and 102 and 103, respectively) and the region between residues 31 and 45. Since the constraints are not evenly distributed throughout the molecule, those regions of the protein that have a larger number of constraints per residue are better defined (Figures 3 and 4).

In conclusion, the resulting structures satisfy the experimental constraints with small deviations from idealized covalent geometry, with most of the backbone torsion angles for non-glycine residues lying within the allowed regions in the Ramachandran plot. The structures show good geometry, a low DYANA function target, a low AMBER energy, and few NOE violations.

Main Chain. The core consists of a globular five-helix motif with a simple "up-and-down" topology, in which the helices are arranged in a right-handed superhelix, a folding motif first observed in nonspecific lipid transfer proteins (nsLTPs) (*45*−*48*), hydrophobic protein from soybean (HPS), and α-amylase inhibitors (*49*−*51*), and subsequently detected in 2S albumins (*18*, *19*). The helices span residues 11−17 (helix Ia), 22−29 (helix Ib), 50−57 (helix II), 65−70 (helix III), and 80−93 (helix IV). Helices Ia and Ib are linked by a nonhelical turn, L1 (residues 18−22). Although Gly10 does not show the conventional angles for an α-helical conformation ($\psi = 162 \pm 63°$, $\phi = -106 \pm 44°$), the hydrogen bond between Glu14(HN) and Gly10(O) is formed in ∼80% of the structures. In contrast, Ala18(HN) has a ψ of $-128 \pm 8°$ and a ϕ of $47 \pm 12°$ and forms a hydrogen bond with

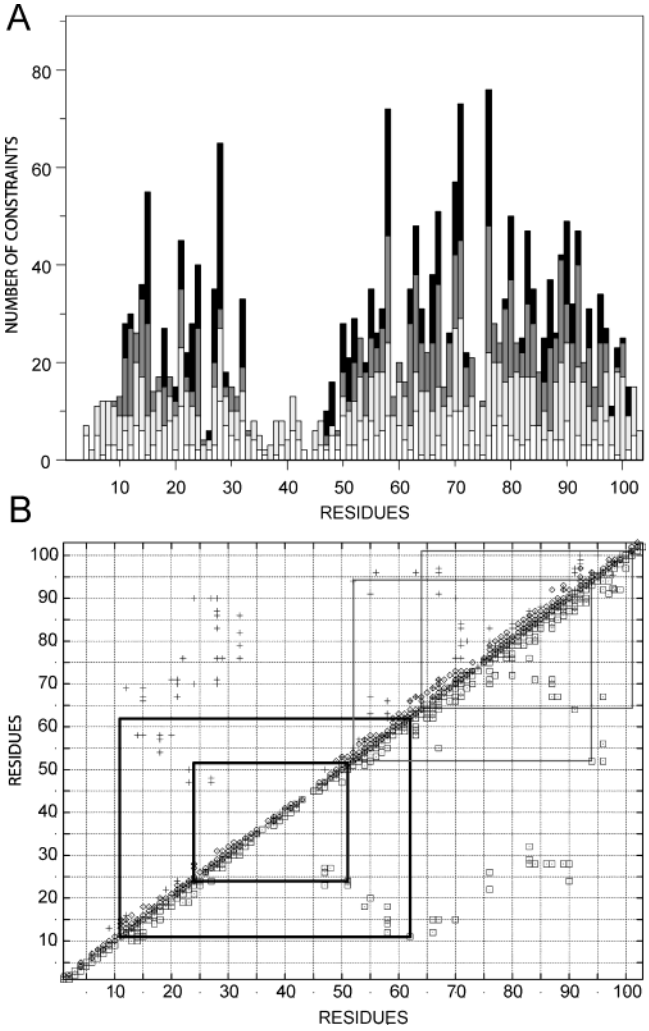

FIGURE 3: (A) Distribution of intraresidual (white), sequential (light gray), short-range (dark gray), and long-range (black) NOEs by residue. (B) Diagonal plot of the interproton NOEs. Above the diagonal are backbone−backbone (◇) and side chain−side chain (+) NOEs. Below the diagonal are side chain−backbone (□) NOEs. Disulfide bridges are linked by lines.

Table 1: Structural Statistics of the 20 Best Structures of SFA-8

Constraints			
distances			
intraresidual			396
sequential			406
medium-range ($i - j < 5$)			335
long-range ($i - j \geq 5$)			256
DYANA			
	average	minimal	maximal
target function	2.33	1.41	3.10
maximal violation (Å)	0.29	0.20	0.39
sum of violations (Å)	9.40	7.80	11.4
AMBER			
mean energy (kcal/mol)			−1161
rms deviations from ideal geometry			
bond lengths (Å)			0.0105
bond angles (deg)			2.520
distance constraint violations			
number >0.4 Å			0
0.4 Å ≥ number > 0.3 Å			2
0.3 Å ≥ number > 0.2 Å			15
maximal violation (Å)			0.35
rmsd radius			
		backbone	heavy atoms
global (residues 11−30 and 46−101)		0.52	1.01
helix Ia		0.23	0.84
helix Ib		0.15	0.49
helix II		0.21	0.65
helix III		0.10	0.52
helix IV		0.28	0.80
Ramachandran			
favorable regions			95.5%
additional allowed regions			4.5%
disallowed regions			0%

Met15(O) in 13 of 20 structures. The average angle formed between these two helices is 75° in the 20 final minimized structures, as compared to 53° in the castor bean albumin rRicC3. A long, unstructured loop region (L2) is present between helices Ib and II, encompassing residues Lys30−Gln49. Helix II runs antiparallel to helices Ia and Ib, while helices Ia and II are linked to the bundle formed by helices III and IV by two disulfide bridges, Cys11 (helix Ia)−Cys62 (helix III) and Cys52 (helix II)−Cys94 (C-terminal segment). The interhelical loop between helices II and III (L3) exhibits a recurrent classical type I β-turn conformation between residues Asp59 and Cys62, with a hydrogen bond between Asp59(HN) and Cys62(O) in all structures. Helices III and IV run almost antiparallel to each other and are connected by a relatively short segment comprising residues Leu71−Met79. This region shows high variability in length and sequence between different members of the 2S albumin family and hence is often termed the "hypervariable region" (*52*−*54*). Residues Leu71 and Asn72 in the C-terminus of helix III have ψ angles of $-95 \pm 5°$ and $-123 \pm 7°$ and ϕ angles of $21 \pm 8°$ and $1 \pm 6°$ and are involved in hydrogen bonds with Met68(O) and Met69(O), respectively, in 50% of the structures. The hypervariable region is organized around the γ-turn (Glu73−Met75), which is stabilized by the hydrogen bond from Met75(HN) to Glu73(O). In the castor bean albumin rRicC3, the hypervariable loop is three residues shorter than in SFA-8 and is constituted by two γ-turns. Residues Ser85, His88, and Asn89 in helix IV of SFA-8 deviate from the helical conformation ($\psi = -75 \pm 11°$, $-102 \pm 3°$, and $-110 \pm 3°$; $\phi = 12 \pm 23°$, $-23 \pm 5°$, and $-22 \pm 3°$), causing helix bending and breaking of the $i-i+4$ hydrogen bond interaction. A similar bending is seen in the analogous helix in nsLTP (*47*). The intersecting angles of the axes among helices II−IV are similar to those observed in rRicC3. The C-terminus of helix IV shows a type I β-turn between residues Ile92 and Asn95, which is stabilized by an Ile92(O)−Asn95(HN) hydrogen bond. Finally, helix IV is followed by a long loop devoid of any regular secondary structure. This packs against helix III, which is favored by the disulfide bridge between Cys64 (L3) and Cys101 (C-terminus). Furthermore, a hydrogen bond between Leu96(HN) and Pro91(O) is present in all the structures.

Two intrahelical proline residues, Pro65 and Pro91, break the regular $i-i+4$ hydrogen bonding pattern of the ideal helical structure in helices III and IV, respectively. Both residues disturb the helical structure without a detectable change in the helix axis direction.

The volume of the protein as calculated with VADAR is 10 951.7 Å^3, a value 20% lower than expected from its molecular weight (14 269.4 Å^3).

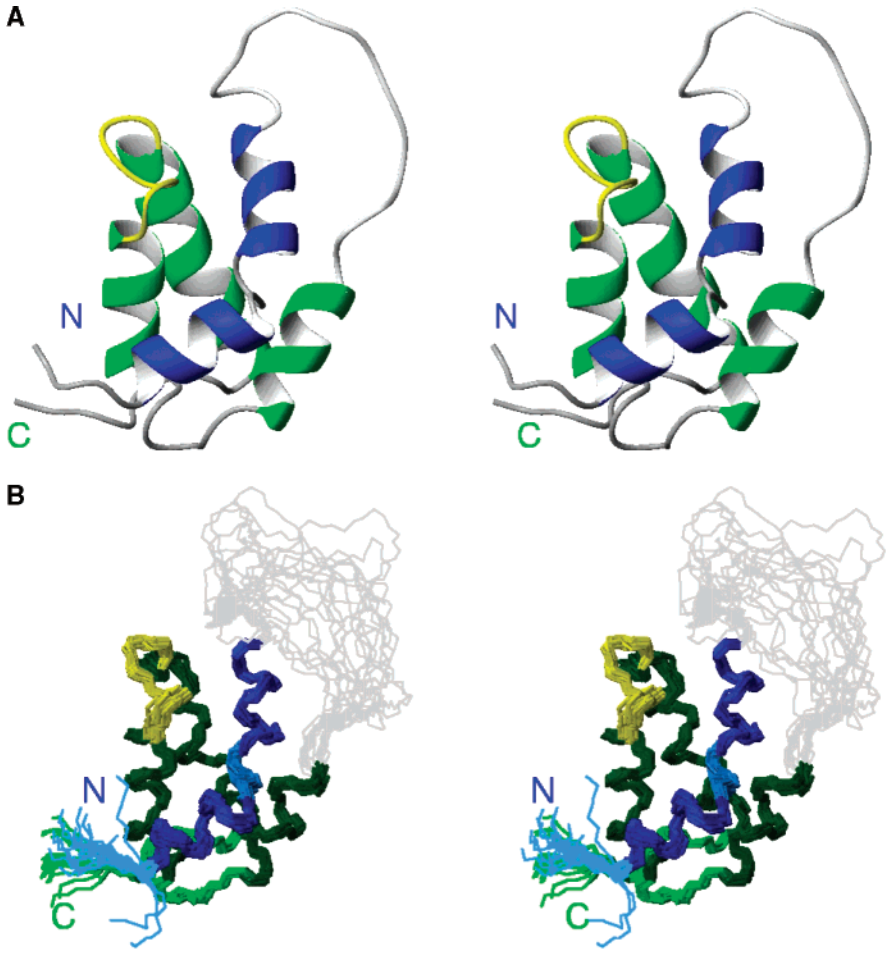

FIGURE 4: (A) Stereoview of a ribbon representation of the solution structure of SFA-8. (B) Stereoview of the superposition of the 20 best DYANA structures of SFA-8 after restrained energy minimization.

Side Chains. Most side chains of the hydrophobic residues, including some alanines, are significantly protected from the solvent and point toward the core of the protein. However, Met26, Val38, Leu50, Met53, Met75, Ile77, Met84, Met86, Cys101, and Met103 have fractional accessible surface areas of >0.4. The side chains of the hydrophobic residues are relatively well packed and well defined conformationally, resulting in small χ_1 circular variances. Met15, Val38, Met42, Met68, Met84, and Met103 display significant variability with circular variances of 0.278, 0.401, 0.233, 0.293, 0.174, and 0.519, respectively, indicating high conformational side chain flexibility. Some clusters of hydrophobic side chains are present in the core of the protein close to the disulfide bonds. The largest cluster is formed by the hydrophobic side chains of helix Ia, the interhelical loop L1 (Ia−Ib), the N-terminus of helix Ib, the C-terminus of helix II, interhelical loop L3 (II−III), the N-terminus of helix III, and the C-terminal loop and includes residues Met15, Met20, Leu21, Leu55, Leu58, Met63, A66, Ile67, and Leu96. Furthermore, three different clusters of hydrophobic residues are present in the region where the side chains of the C-termini of helices Ib and III and the N-terminus of helix IV interact with the hypervariable loop. One of these involves residues Met26, Met29, Met75, and Met79, a second one Leu28, Leu32, Met86, and Leu90, and the last one Met70, Leu71, Met84, and Ala87. Such hydrophobic associations are close to the protein surface.

SFA-8 has 13 negatively (Asp and Glu) and 14 positively (Arg, His, and Lys) charged residues. Most of these charged residues are located on the surface of the protein and show high fractional accessibility values. The main exceptions are His23 and His47, which are packed close to each other, and Glu73 and Arg80 in the hypervariable loop which show accessibility values of less than 0.35. The distribution of charged residues in the protein surface is not uniform with a cluster of negative charges being located at the confluence of the C-terminus of helix I (Glu16, Glu17, and Glu19) and the C-terminus of helix III (Glu73).

DISCUSSION

Comparison with Structurally Related Proteins. CLUSTALW (*55*) was used to perform a multiple-sequence alignment of different 2S albumin proteins, including the napin BnIb and the castor bean albumin rRicC3. The resulting alignment showed a low level of sequence homology with even the cysteine positions being imperfectly aligned, despite the fact that all proteins share the same pattern of cysteines and the same arrangement of disulfide bonds. The use of more sensitive algorithms for multiple-sequence alignments and hidden Markov models such as Pfam (*56*) showed that the SFA-8 sequence belongs to the protease inhibitor/seed storage/LTP superfamily with an *E* value of 8.2×10^{-11}. Four families constitute this superfamily: 2S albumin seed storage proteins, nonspecific lipid transfer proteins (nsLTPs), α-amylase inhibitors, and the hydrophobic protein from

Table 2: Summary of Structural Alignment Statistics

	Z score for DALI	rmsd (Å)	aligned Cα/total Cα	sequence identity (%)
α-amylase inhibitor from maize	5.8	2.9	82/116	24
2S albumin from *R. communis* (rRicC3)	4.3	2.9	78/125	21
nsLTP from maize	3.5	2.8	69/93	7
HPS protein	3.1	3.4	60/75	13

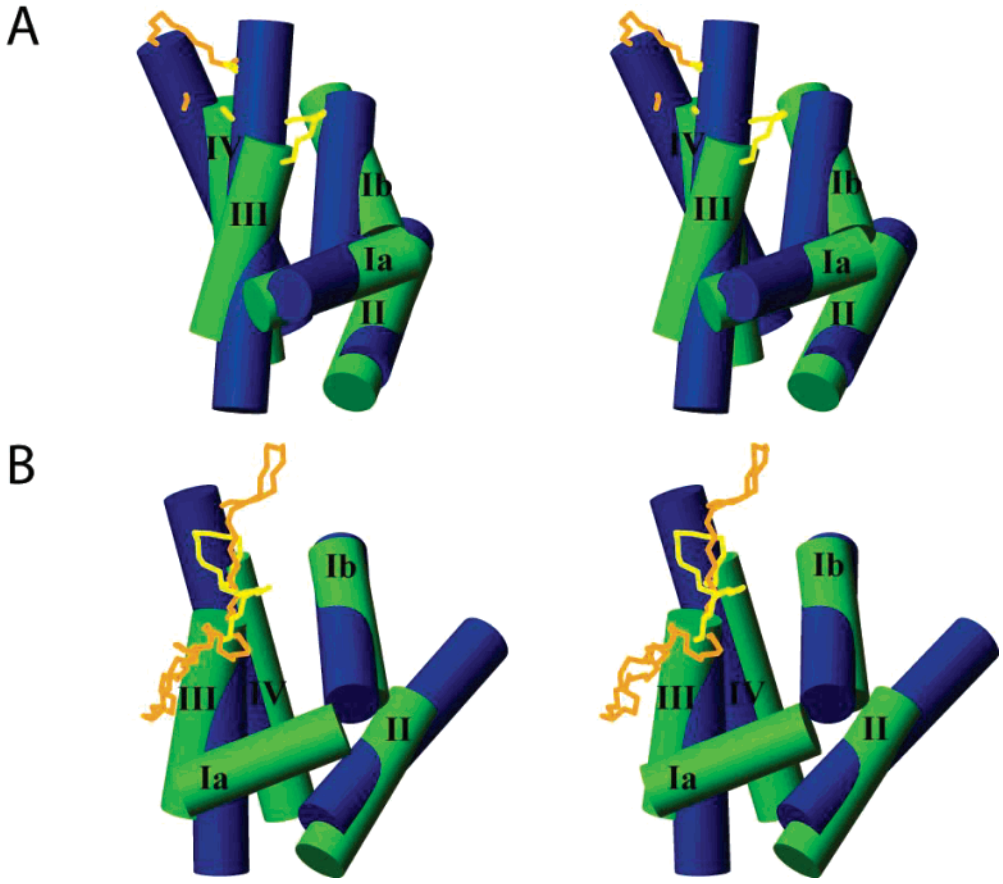

FIGURE 5: (A) Superposition of the solution structures of SFA-8 (1S6D, in green; hypervariable loop in yellow) and RicC3 (1PSY, in blue; hypervariable loop, in orange). (B) Superposition of the solution structures of SFA-8 (1S6D, in green; hypervariable loop, in yellow), and the α-amylase inhibitor from maize (1BFA, in blue; fingerloop, in orange).

soybean (HPS). The fold corresponding to this superfamily is described in the Structural Classification of Protein database [SCOP (*20*)] as "4 helices; folded leaf; right-handed superhelix; disulfide rich". Furthermore, when the whole Protein Data Bank was screened with the 3D structure of SFA-8 for fold recognition using the Dali server (*57*), the highest score was obtained with this superfamily. A summary of the structural alignment statistics of rRicC3 [1PSY (*19*)], the α-amylase inhibitor from maize [1BFA (*49*)], the nsLTP from maize [1MZM (*58*)], and the HPS protein [1HYP (*59*)] is presented in Table 2. According to these data, the 3D structure of SFA-8 is more similar to those of α-amylase inhibitors from maize than to those of other 2S albumins. This is also in agreement with previous modeling studies (*25*).

SFA-8 and the recombinant RicC3 have very similar 3D structures (Figure 5A). Furthermore, in both 3D structures, the linker region is unstructured and does not interact with any other part of the protein, indicating that it does not contribute to the stability of this family of proteins. The most important difference between the two proteins is in the arrangement of the hypervariable loop and the N-terminus of helix IV. This region is much closer to helix Ib in SFA-8 than in RicC3. Furthermore, helix III is 11 residues shorter than in rRicC3, while the hypervariable loop is three residues longer. According to the *Z* score of the Dali server, SFA-8 is more similar to α-amylase inhibitors than to rRicC3 in the arrangement of the α-helices, though helix Ia is not present in α-amylase inhibitors and the lengths and conformations of the segments connecting the α-helices differ considerably (Figure 5B). A special mention has to be made of the hypervariable region in 2S albumins, which is an extended loop of nine residues in SFA-8, but comprises a protruding segment of 22 residues with partial *β*-structure in the α-amylase inhibitor of maize. This loop is also present in the α-amylase inhibitor from ragi (Indian finger millet) (*51*, *60*) and is shifted toward the C-terminus in the α-amylase inhibitor from wheat (*50*).

The pattern of disulfide bonds of SFA-8 is identical to that in rRicC3 and α-amylase inhibitors but differs from that in nsLTPs. As in rRicC3, the different orientation of helix III in SFA-8 compared with that in nsLTPs, by ∼20°, precludes the formation of an internal cavity (*19*), and hence binding of lipids. In the case of SFA-8, the absence of an internal cavity is demonstrated by detailed examination of the structures as well as application of VOIDOO (*61*). This

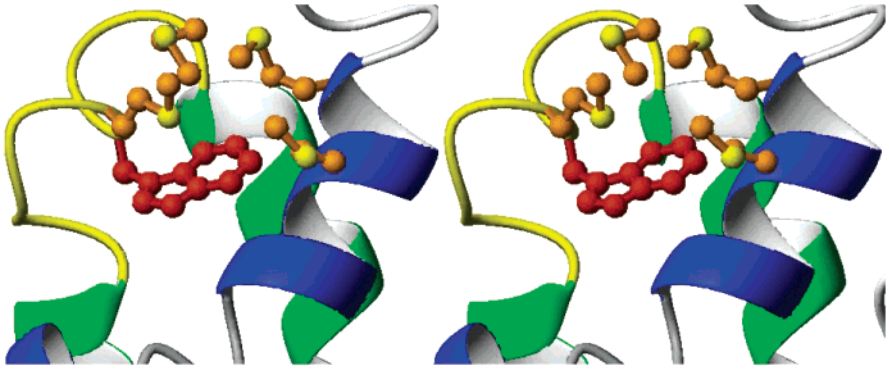

FIGURE 6: Stereoview of the hydrophobic patch involving Trp76 (in red) and Met26, Met29, Met75, and Met79 (in orange).

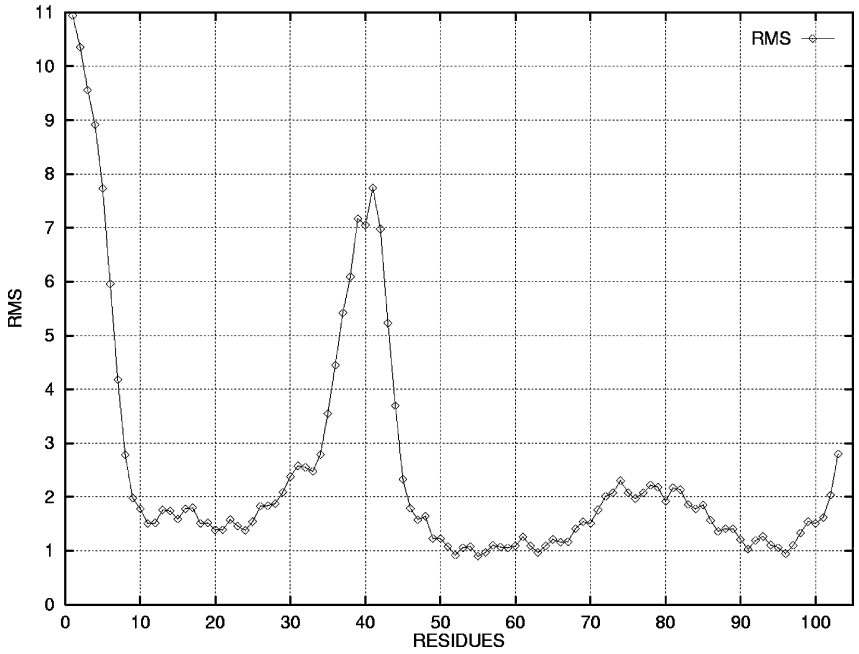

FIGURE 7: Atomic rms deviations in angstroms per residue of a geometrical average of the 20 best structure estimates of SFA-8 for backbone atoms (N, Cα, C, and O).

observation is consistent with the hypothesis that the pattern of cysteine residue pairing determines whether a hydrophobic cavity can be formed (*19*).

Role of Trp76 and Hydrophobic Patches in Determining Emulsification Properties. SFA-8 has been shown to be highly active in stabilizing oil/water emulsions (*21*). SFA-8 is strongly retained on RP-HPLC, which may be explained by the presence of a cluster of hydrophobic residues on the protein surface. Furthermore, the presence of a hydrophobic face on the surface of SFA-8 might be an important factor in stabilizing emulsions by providing a mechanism for the initial adsorption of the protein to a water−oil or water−air interface through hydrophobic interactions (*26*).

QUILT (*62*) was therefore used to detect hydrophobic patches on the surface of SFA-8, using the 20 final minimized structures. This identified four hydrophobic regions with solvent-accessible surface areas of 250−500 Å^2. One patch is located at the confluence of the C-terminus of helix Ib and the N-terminus of helices II and IV (Tyr27, Leu28, Leu32, Met86, and Leu90). A second one involves the interhelical space between helix III, helix IV, and the C-terminus (Ile67, Met68, Met69, Leu71, Met84, Ile92, and Met97). A third one is located in the external face of helix II and the N-terminus of helix Ib (His23, His47, Leu50, and Met53). Finally, the largest one, containing the hypervariable loop and helix Ib, is located above and partially covering Trp76 (Met26, Met29, Met75, Ile77, and Met79).

Fluorescence studies of SFA-8 in solution (*26*) show that Trp76 (emission peak maximum $\lambda_{em} = 340$ nm) is a class II tryptophan, being exposed to water to a limited extent. This conclusion is in general agreement with the solution structure of SFA-8 determined herewith. Under the NMR conditions, the side chain of Trp76 has a fractional accessible surface area of ~0.06, with the indole group located at the interior of the protein and oriented in a plane parallel to the protein surface. Furthermore, hydrophobic residues Met26, Met29, Met75, and Met79 are packed together, forming a "crown" above the indole group of Trp76 (Figure 6). In contrast, the fluorescence spectrum of SFA-8 in an emulsion of 20% (v/v) *n*-hexadecane with an aqueous SFA-8 solution has a peak maximum at 312 nm. This pronounced blue shift of the emission peak is characteristic of the Trp76 chromophore moving into a more nonpolar environment, i.e., it has probably become located in the oil phase following adsorption. The Trp residue is located at the center of a segment with no regular secondary structure known as the hyper-

varible region in 2S albumins. That segment exhibits the highest values for the atomic backbone rmsd of the protein with the obvious exception of the 10 N-terminal residues and those in the segment of residues 30−50 (see Figure 7). The estimation of the S^2 order parameters of the N−H vectors of the backbone (*63*), which can be considered as a measure of local protein flexibility, indicates that the hypervariable region in SFA-8 shows the highest mobility among all structural elements (data not shown), in agreement with rmsd data. In addition, the hydrophobic region involving the Trp residue is the one with the largest surface area (∼500 Å^2). This combination of flexibility and hydrophobicity may well facilitate a conformational change in this region of the protein when an emulsion is formed, with the Trp76 side chain and probably some of the neighboring methionines reorienting to be directed toward the oil phase. Hydrophobic and charged areas are segregated on the surface of SFA-8. Therefore, if the largest hydrophobic patch containing Trp76 interacts directly with the oil droplet, the remaining areas may be available for protein−protein interactions, accounting for the large increase in the dilational elastic modulus, which is exhibited by a film of SFA-8 as a function of the surface pressure (*26*).

Does the Hypervariable Region Relate to Allergenicity? Although 2S albumins are major allergens in a number of plant seeds and nuts (including mustards, walnut, Brazil nut, castor bean, and peanut) (*2*, *64*), it is not yet clear whether SFA-8 is an allergen. Several American patients who are allergic to sunflower seeds have been reported to have IgE that reacts with 2S albumin proteins from sunflower (*23*, *24*). However, despite the high level of consumption of sunflower seeds in Southern Europe, the incidence of allergy is very low, and recent studies of a group of Greek patients have demonstrated reactions with the nsLTP but not SFA-8 or other 2S albumins (*65*). It has been noted that allergenic proteins are often resistant to proteolytic digestion, and it has been suggested that this may facilitate their persistence into the small intestine (*66*, *67*). It has therefore been suggested that resistance to proteolysis can be used as an indicator for the potential of novel proteins to act as allergens, and included in protocols for use in assessing novel and genetically modified foods. The 2S albumins are indeed highly resistant to proteolysis, as concluded in a recent study showing that SFA-8 is stable to pepsin treatment (*68*), which is in agreement with the compactness of its 3D structure.

Studies of allergen epitopes have also suggested that sequences present in the hypervariable loop regions of 2S albumins and related proteins may be important in determining allergenicity (*2*, *64*). Thus, the major 2S albumin allergen of yellow mustard (Sin a 1) contains two immunodominant epitopes, one of which is located in the hypervariable region (*16*, *69*). Similarly, the 2S albumin allergen of English walnut (Jug r 1) contains a single immunodominant epitope, which has been identified as a tetrapeptide (RGEE) in the hypervariable region (*70*).

Residues Asn72−Met79 constitute the hypervariable region in SFA-8, with the side chains of Asn72, Pro74, Met75, Ile77, and Arg78 having a large fractional accessible surface area, but with Glu73, Trp76, and Met79 being buried. This contrasts with rRicC3, in which all residues in the hypervariable loop have large fractional accessible surface areas. Thus, the average value of the accessible surface area for the residues in the hypervariable loop of SFA-8 is 0.37, versus a value of 0.64 in rRicC3. Also, the His96-Gly97-Glu98-Glu99 segment in rRicC3 is directed away from the protein surface. These conformational differences between the two 2S albumin proteins may account for the fact that SFA-8 has a lower allergenicity than rRicC3, since access of IgE antibodies to the hypervariable loop is much easier in RicC3 than in SFA-8.

ACKNOWLEDGMENT

We are grateful to Mrs. C. López for technical assistance and Dr. Roger Fido, Dr. Alison Lovegrove, and Johnathan Griffin (Rothamsted Research) for assisting with protein purification and for carrying out MS analyses.

REFERENCES

1. Shewry, P. R., and Pandya, M. J. (1999) in *Seeds Proteins* (Shewry, P. R., and Casey, R., Eds.) Kluwer Academic Publishers, Dordrecht, The Netherlands.
2. Monsalve, R. I., Villalba, M., Rico, M., Shewry, P. R., and Rodríguez, R. (2003) The 2S albumin proteins, in *Plant Food Allergens* (Mills, E. N. C., and Shewry, P. R., Eds.) pp 42−56, Blackwell Science, Oxford, U.K.
3. Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, DC.
4. Kreis, M., Forde, B. G., Rahman, S., Miflin, B. J., and Shewry, P. R. (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat, *J. Mol. Biol. 183*, 499−502.
5. Terras, F., Schoofs, H., Thevissen, K., Osborn, R. W., Vanderleyden, J., Cammue, B., and Broekaert, W. F. (1993) Synergistic enhancement of the antifungal activity of wheat and barley thionins by radish and oilseed rape 2S albumins and by barley trypsin inhibitors, *Plant Physiol. 103*, 1311−1319.
6. Terras, F. R., Schoofs, H. M., De Bolle, M. F., Van Leuven, F., Rees, S. B., Vanderleyden, J., Cammue, B. P., and Broekaert, W. F. (1992) Analysis of two novel classes of plant antifungal proteins from radish (*Raphanus sativus* L.) seeds, *J. Biol. Chem. 267*, 15301−15309.
7. Terras, F. R., Torrekens, S., Van Leuven, F., Osborn, R. W., Vanderleyden, J., Cammue, B. P., and Broekaert, W. F. (1993) A new family of basic cysteine-rich plant antifungal proteins from Brassicaceae species, *FEBS Lett. 316*, 233−240.
8. Genov, N., Goshev, I., Nikolova, D., Georgieva, D. N., Filippi, B., and Svendsen, I. (1997) A novel thermostable inhibitor of trypsin and subtilisin from the seeds of *Brassica nigra*: amino acid sequence, inhibitory and spectroscopic properties and thermostability, *Biochim. Biophys. Acta 1341*, 157−164.
9. Svendsen, I., Nicolova, D., Goshev, I., and Genov, N. (1994) Primary structure, spectroscopic and inhibitory properties of a two-chain trypsin inhibitor from the seeds of charlock (*Sinapis arvensis* L), a member of the napin protein family, *Int. J. Pept. Protein Res. 43*, 425−430.
10. Svendsen, I. B., Nicolova, D., Goshev, I., and Genov, N. (1989) Isolation and characterization of a trypsin inhibitor from the seeds of kohlrabi (*Brassica napus* var. rapifera) belonging to the napin family of storage proteins, *Carlsberg Res. Commun. 54*, 231−239.
11. Vandekerckhove, J., van Damme, J., van Lijsebettens, M., Botterman, J., De Block, M., Vandewiele, M., De Clercq, A., Leemans, J., van Montagu, M., and Krebbers, E. (1989) Enkephalins produced in transgenic plants using modified 2S seed storage proteins, *Bio/Technology 7*, 929−932.
12. Altenbach, S. B., Kuo, C. C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A., and Townsend, J. (1992) Accumulation of a brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine, *Plant Mol. Biol. 18*, 235−245.
13. Wolff, N., Cogan, U., Admon, A., Dalal, I., Katz, Y., Hodos, N., Karin, N., and Yannai, S. (2003) Allergy to sesame in humans is associated primarily with IgE antibody to a 14 kDa 2S albumin precursor, *Food Chem. Toxicol. 41*, 1165−1174.

14. Pastorello, E. A., Varin, E., Farioli, L., Pravettoni, V., Ortolani, C., Trambaioli, C., Fortunato, D., Giuffrida, M. G., Rivolta, F., Robino, A., Calamari, A. M., Lacava, L., and Conti, A. (2001) The major allergen of sesame seeds (*Sesamum indicum*) is a 2S albumin, *J. Chromatogr., B: Biomed. Sci. Appl. 756*, 85−93.
15. Thorpe, S. C., Kemeny, D. M., Panzani, R., McGurl, B., and Lord, M. (1998) Allergy to castor bean: II. Identification of the major allergens in castro bean seeds, *J. Allergy Clin. Immunol. 82*, 67−72.
16. Menéndez-Arias, L., Domínguez, J., Moneo, I., and Rodríguez, R. (1990) Epitope mapping of the major allergen from yellow mustard (*Sinapis alba* L.), *Mol. Immunol. 27*, 143−150.
17. Molvig, L., Tabe, L. M., Eggum, B. O., Moore, A. E., Craig, S., Spencer, D., and Higgins, T. J. (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene, *Proc. Natl. Acad. Sci. U.S.A. 94*, 8393−8398.
18. Rico, M., Bruix, M., Gonzalez, C., Monsalve, R. I., and Rodriguez, R. (1996) ^{1}H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein, *Biochemistry 35*, 15672−15682.
19. Pantoja-Uceda, D., Bruix, M., Giménez-Gallego, G., Rico, M., and Santoro, J. (2003) Solution structure of RicC3, a 2S albumin storage protein from *Ricinus communis*, *Biochemistry 42*, 13839−13847.
20. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol. 247*, 536−540.
21. Guéguen, J., Popineau, Y., Anisimova, I. N., Fido, R. J., Shewry, P. R., and Tatham, A. S. (1996) Functionality of the 2S albumin seed storage proteins from sunflower (*Helianthus annuus* L.), *J. Agric. Food Chem. 44*, 1184−1189.
22. Kortt, A. A., Caldwell, J. B., Lilley, G. G., and Higgins, T. J. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.), *Eur. J. Biochem. 195*, 329−334.
23. Kelly, J. D., Hlywka, J. J., and Hefle, S. L. (2000) Identification of sunflower seed IgE-binding proteins, *Int. Arch. Allergy Immunol. 121*, 19−24.
24. Kelly, J. D., and Hefle, S. L. (2000) 2S methionine-rich protein (SSA) from sunflower seed is an IgE-binding protein, *Allergy 55*, 556−560.
25. Pandya, M. J., Sessions, R. B., Williams, P. B., Dempsey, C. E., Tatham, A. S., Shewry, P. R., and Clarke, A. R. (2000) Structural characterization of a methionine-rich, emulsifying protein from sunflower seed, *Proteins 38*, 341−349.
26. Burnett, G. R., Rigby, N. M., Mills, E. N. C., Belton, P. S., Fido, R. J., Tatham, A. S., and Shewry, P. R. (2002) Characterization of the emulsification properties of 2S albumins from sunflower seed, *J. Colloid Interface Sci. 247*, 177−185.
27. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR 2*, 661−665.
28. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego.
29. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J. Biomol. NMR 5*, 1−10.
30. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
31. Seavey, B. R., Farr, E. A., Westler, W. M., and Markley, J. L. (1991) A relational database for sequence-specific protein NMR data, *J. Biomol. NMR 1*, 217−236.
32. Güntert, P., Braun, W., and Wüthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA, *J. Mol. Biol. 217*, 517−530.
33. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
34. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
35. Mumenthaler, C., Güntert, P., Braun, W., and Wüthrich, K. (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination, *J. Biomol. NMR 10*, 351−362.
36. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J. Mol. Biol. 319*, 209−227.
37. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Wang, J., Ross, W. S., Simmerling, C., Darden, T., Merz, K. M., Stanton, R. V., Cheng, A., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P., and Kollman, P. A. (2002) *AMBER 7*, University of California, San Francisco.
38. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
39. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
40. Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry, *J. Mol. Biol. 182*, 295−315.
41. Alcocer, M. J., Murtagh, G. J., Bailey, K., Dumoulin, M., Meseguer, A. S., Parker, M. J., and Archer, D. B. (2002) The disulphide mapping, folding and characterisation of recombinant Ber e 1, an allergenic protein, and SFA8, two sulphur-rich 2S plant albumins, *J. Mol. Biol. 324*, 165−175.
42. Egorov, T. A., Odintsova, T. I., Musolyamov, A., Fido, R., Tatham, A. S., and Shewry, P. R. (1996) Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily, *FEBS Lett. 396*, 285−288.
43. Bundi, A., and Wüthrich, K. (1979) ^{1}H NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH, *Biopolymers 18*, 285−297.
44. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR 5*, 67−81.
45. Gincel, E., Simorre, J. P., Caille, A., Marion, D., Ptak, M., and Vovelle, F. (1994) Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional ^{1}H NMR data. A new folding for lipid carriers, *Eur. J. Biochem. 226*, 413−422.
46. Gomar, J., Petit, M. C., Sodano, P., Sy, D., Marion, D., Kader, J. C., Vovelle, F., and Ptak, M. (1996) Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds, *Protein Sci. 5*, 565−577.
47. Heinemann, B., Andersen, K. V., Nielsen, P. R., Bech, L. M., and Poulsen, F. M. (1996) Structure in solution of a four-helix lipid binding protein, *Protein Sci. 5*, 13−23.
48. Poznanski, J., Sodano, P., Suh, S. W., Lee, J. Y., Ptak, M., and Vovelle, F. (1999) Solution structure of a lipid transfer protein extracted from rice seeds. Comparison with homologous proteins, *Eur. J. Biochem. 259*, 692−708.
49. Behnke, C. A., Yee, V. C., Trong, I. L., Pedersen, L. C., Stenkamp, R. E., Kim, S. S., Reeck, G. R., and Teller, D. C. (1998) Structural determinants of the bifunctional corn Hageman factor inhibitor: X-ray crystal structure at 1.95 Å resolution, *Biochemistry 37*, 15277−15288.
50. Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T., and Morimoto, T. (1997) Tertiary and quaternary structures of 0.19 α-amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 Å resolution, *Biochemistry 36*, 13503−13511.
51. Strobl, S., Muhlhahn, P., Bernstein, R., Wiltscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R., and Holak, T. A. (1995) Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin inhibitor from ragi seeds by NMR spectroscopy, *Biochemistry 34*, 8281−8293.
52. Raynal, M., Depigny, D., Grellet, F., and Delseny, M. (1991) Characterization and evolution of napin-encoding genes in radish and related crucifers, *Gene 99*, 77−86.
53. Monsalve, R. I., López-Otín, C., Villalba, M., and Rodríguez, R. (1991) A new distinct group of 2S albumins from rapeseed. Amino acid sequence of two low molecular weight napins, *FEBS Lett. 295*, 207−210.
54. Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and VandeKerckhove, J. (1988) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family, *Plant Physiol. 87*, 859−866.

55. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
56. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database, *Nucleic Acids Res. 30*, 276−280.
57. Holm, L., and Sander, C. (1996) Mapping the protein universe, *Science 273*, 595−603.
58. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K., and Suh, S. W. (1995) High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings, *Structure 3*, 189−199.
59. Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S., and Lehmann, M. S. (1993) Crystal structure of hydrophobic protein from soybean: a member of a new cysteine-rich family, *J. Mol. Biol. 231*, 877−887.
60. Gourinath, S., Alam, N., Srinivasan, A., Betzel, C., and Singh, T. P. (2000) Structure of the bifunctional inhibitor of trypsin and α-amylase from ragi seeds at 2.2 Å resolution, *Acta Crystallogr. D56*, 287−293.
61. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures, *Acta Crystallogr. D50*, 178−185.
62. Lijnzaad, P., Berendsen, H. J., and Argos, P. (1996) Hydrophobic patches on the surfaces of protein structures, *Proteins 25*, 389−397.
63. Zhang, F., and Bruschweiler, R. (2002) Contact Model for the Prediction of NMR N−H Order Parameters in Globular Proteins, *J. Am. Chem. Soc. 124*, 12654−12655.
64. Monsalve, R. I., Villalba, M., and Rodríguez, R. (2001) Allergy to mustard seeds: The importance of 2S albumins as food allergens, *Internet Symp. Food Allergens 3*, 57−69.
65. Papageorgiou, P. S., Shewry, P. R., Fido, R. J., Papadopoulos, N. G., Tassios, I. K., and Tatham, A. S. (2003) Identification of lipid transfer protein (LTP) as a major sunflower seed allergen in sunflower seed-induced systemic reactions, *Proceedings of the Greek National Meeting of Allergy and Clinical Immnunology, Athens*.
66. Astwood, J. D., Leach, J. N., and Fuchs, R. L. (1996) Stability of food allergens to digestion in vitro, *Nat. Biotechnol. 14*, 1269−1273.
67. Mills, E. N. C., Jenkins, J. A., and Shewry, P. R. (2003) The role of common properties in determining plant food protein allergenicity, in *Plant Food Allergens* (Mills, E. N. C., and Shewry, P. R., Eds.) pp 158−170, Blackwell Science, Oxford, U.K.
68. Murtagh, G. J., Archer, D. B., Dumoulin, M., Ridout, S., Matthews, S., Arshad, S. H., and Alcocer, M. J. (2003) In vitro stability and immunoreactivity of the native and recombinant plant food 2S albumins Ber e 1 and SFA-8, *Clin. Exp. Allergy 33*, 1147−1152.
69. Monsalve, R. I., González de la Peña, M., López-Otín, C., Villalba, M., and Rodríguez, R. (1993) Characterization of a new oriental-mustard (*Brassica juncea*), Braj IE: detection of an allergenic epitope, *Biochem. J. 293*, 625−632.
70. Robotham, J. M., Suzanne, S., Teuber, S., Sathe, S. K., and Roux, K. H. (2002) Linear IgE epitope mapping of the english walnut (*Junglas regia*) major food allergen, Jug r 1, *J. Allergy Clin. Immunol. 109*, 143−149.

BI0496900