Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Purification, identification and preliminary crystallographic studies of a 2S albumin seed protein from *Lens culinaris*

Pankaj Gupta, Vineet Gaur and Dinakar M. Salunke*

National Institute of Immunology, New Delhi 110 067, India

Correspondence e-mail: dinakar@nii.res.in

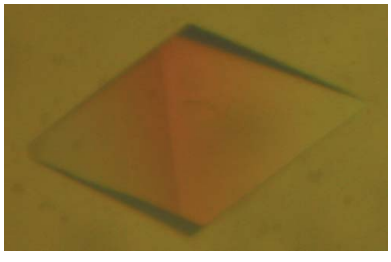
Received 28 March 2008
Accepted 14 July 2008

Lens culinaris (lentil) is a widely consumed high-protein-content leguminous crop. A 2S albumin protein (26.5 kDa) has been identified using NH_2-terminal sequencing from a 90% ammonium sulfate saturation fraction of total *L. culinaris* seed protein extract. The NH_2-terminal sequence shows very high homology to PA2, an allergy-related protein from *Pisum sativum*. The 2S albumin protein was purified using a combination of size-exclusion and ion-exchange chromatography. Crystals of the 2S seed albumin obtained using the hanging-drop vapour-diffusion method diffracted to 2.5 Å resolution and were indexed in space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 78.6$, $c = 135.2$ Å.

1. Introduction

Allergy is a type I hypersensitivity reaction that elicits specific IgE antibodies against innocuous antigens which bind to high-affinity Fc receptors on mast cells and basophils, resulting in the release of biological mediators (*e.g.* histamine, leukotriene) responsible for the effecter phase of the hypersensitivity reaction. Food allergies are among the most common causes of hypersensitivity, being second only to pollen allergies (Madsen, 2005). Plant food allergies are a major concern and affect almost 4% of the human population all over the world (Dall'Antonia *et al.*, 2005). Food hypersensitivity is an immunological reaction arising from the ingestion of food in certain susceptible individuals, resulting in symptoms ranging from respiratory and gastrointestinal allergy to cardiovascular anaphylaxis and sometimes systemic anaphylactic reactions. Allergies to milk, eggs, peanuts, soya and wheat usually resolve by the age of 5–6 y with the maturation of gut mucosal barriers. However, allergies to tree nuts and legumes persist throughout life (Breiteneder & Radauer, 2004). Genetic susceptibility, food-processing methods and hygiene standards generally influence the prevalence of food allergies in a given population (Bischoff & Crowe, 2005). Various biochemical and biophysical characteristics including pH stability, thermal stability, resistance to proteolysis, ligand interactions *etc.* have been associated with the allergenicity of proteins (Breiteneder & Mills, 2005). However, none of these properties provide a satisfactory explanation for the allergenicity of certain proteins over others. A comparative structural analysis of allergy-related proteins should provide better insights into the molecular basis of food allergy, which would help in the design of diagnostic and therapeutic approaches. Also, such studies will shed light on the physiological roles of these proteins in plants.

Most of the plant food allergens belong to three protein superfamilies: the cupins, prolamins and pathogen-related proteins (Breiteneder & Radauer, 2004). The 2S albumins, which are a major component of storage proteins, are low-molecular-weight cysteine-rich proteins that are present in both monocotyledonous and dicotyledonous seeds and belong to the prolamin superfamily. These proteins are normally stable to heat and proteolysis. The 2S albumins are synthesized in seeds at a specific time and play a vital role in the storage of nitrogen and carbon as well as having antifungal properties



© 2008 International Union of Crystallography
All rights reserved

(Breiteneder & Radauer, 2004). Several tree-nut and seed allergens are 2S albumins, including Ber e 1 from *Bertholletia excelsa*, Jug r 1 from *Juglans regia*, Ana o 3 from *Anacardium occidentale*, Ses i 2 from *Sesamum indicum*, Sin a 1 from *Brassica alba* seeds and Bra j 1 from *Brassica juncea* seeds (Breiteneder & Radauer, 2004).

Pulses belong to the Leguminoseae family. They are a rich source of proteins, many of which are associated with immediate hypersensitive responses. *Lens culinaris* (lentil or masoor) is a bushy annual leguminous crop grown for its lens-shaped seeds. The plant originated in the Near East and has been part of the human diet since the aceramic Neolithic period. *L. culinaris* contains 26% protein, which is more then any other pulse with the exception of soybeans. *L. culinaris* is the most common legume involved in allergic reactions in many parts of the world, especially the Mediterranean area, South Asia and India. Allergic manifestations arising from *L. culinaris* ingestion involve various systemic symptoms including cutaneous reactions, oropharyngeal symptoms, acute urticaria or anaphylaxis (López-Torrejón *et al.*, 2003).

Our approach is to investigate the role of plant seed proteins in allergenicity by sequence-based identification of allergenic proteins and their structural analyses by X-ray crystallography in order to establish a molecular basis for allergenicity. Towards this end, we have previously described the purification and crystallization of novel seed proteins from *Vigna unguiculata* (Chanana *et al.*, 2004), *Lathyrus sativus* (Qureshi *et al.*, 2006) and *Prunus dulcis* (Gaur *et al.*, 2008). Here, we report the preliminary crystallographic studies of a 2S albumin from *Lens culinaris*.

2. Experimental procedures

2.1. Protein purification

L. culinaris mature seeds were purchased from the NSC, IARI, New Delhi. 250 g of seeds were ground to a fine powder and defatted with petroleum ether. The protein was extracted at 277 K by resuspending the defatted powder in Milli-Q water containing plant protease-inhibitor cocktail (Sigma, P-9599) by continuous stirring for 4 h followed by centrifugation at 16 799*g* for 90 min. Proteins were partially separated on the basis of their differential solubilities by ammonium sulfate saturation fractionation, which was carried out by subjecting the supernatant to increasing concentrations of ammonium sulfate. The protein pellets thus obtained were resuspended and dialyzed in 50 m*M* Tris–HCl pH 7.5 containing 150 m*M* NaCl and analysed on SDS–PAGE (Fig. 1). Protein obtained in the 90% fraction was subjected to purification by size-exclusion chromatography (SEC). Sephacryl-300 matrix (Pharmacia) was used for SEC in a manually packed column (XK 26 mm/100 cm, Pharmacia). The column was pre-equilibrated with 50 m*M* Tris–HCl pH 7.5 containing 150 m*M* NaCl at a flow rate of 1 ml min^{-1} (Fig. 2*a*). The purified protein was concentrated to 12 mg ml^{-1} using ultrafiltration (Amicon, 10 kDa cutoff, Millipore) and dialyzed against 50 m*M* Tris–HCl pH 8.0. Further purification was achieved using a POROS PI/M weak anion-exchange prepacked column (10 × 100 mm, Applied Biosystems). This PI/M weak anion-exchange column was packed with matrix containing polyethyleneimine as a functional group. Elution was carried out with an isocratic gradient of 0–1 *M* NaCl in 50 m*M* Tris–HCl pH 8.0 over a period of 30 min with a flow rate of 5 ml min^{-1} (25.3 × the bed volume). Albumin was eluted at 160 m*M* NaCl concentration (Fig. 2*b*). The protein concentration was estimated by a BCA protein assay (Pierce Biotechnology) using bovine serum albumin as the standard. A concentration of 6 mg ml^{-1} was achieved using a Centriplus YM10 device (Millipore).

2.2. Protein-sequence analysis

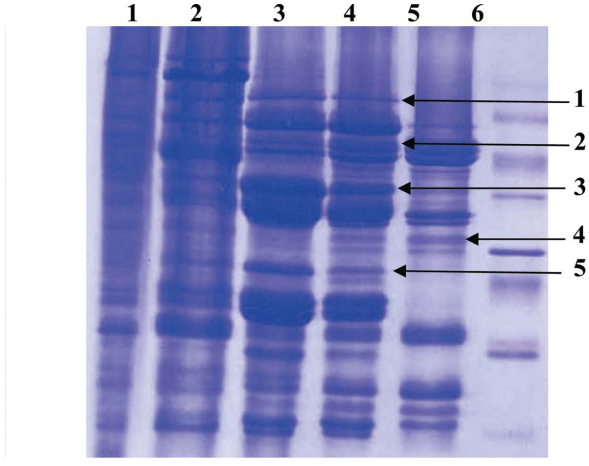
The protein bands were transferred onto a polyvinylidene fluoride (PVDF) membrane using *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) buffer. NH_2-terminal sequencing was carried out using the Edman degradation method on a Procise protein sequencer (Applied Biosystems). The NH_2-terminal sequences were subjected to a *BLAST* homology search to identify proteins with significant sequence homology in a nonredundant database of protein sequences (http://www.ncbi.nlm.nih.gov/BLAST).

2.3. Crystallization

Crystallization trials were performed at room temperature using the hanging-drop vapour-diffusion method (McPherson, 1982). Initial crystallization experiments were set up using Crystal Screens I and II (Hampton Research) as well as many other independently designed conditions. Additionally, pH, temperature and precipitant variations were also tried. More than 300 drops were initially set up for crystallization trials. Drops containing equal volumes (3 µl) of protein solution (6 mg ml^{-1}) and precipitant solution were equilibrated against 1000 µl reservoir solution. Poorly diffracting crystals were obtained in 100 m*M* Tris–HCl pH 7.5 containing 2.4 *M* ammonium sulfate and 10% PEG 600. The final crystallization conditions were optimized to 0.1 *M* MES pH 6.5 containing 10%(*w*/*v*) PEG 1000 and 2 *M* ammonium sulfate.

2.4. Data collection and processing

X-ray diffraction data were collected at 120 K using 33%(*w*/*v*) xylitol (CryoPro kit, HR2-132, Hampton Research) as a cryopro-



Band 1	**NDEEGSELRVPLQRERNRQE**
Band 2	**SRSDENNPFIFESNRFQTLF**
Band 3	**TRSEFDRLNQCQLDNINALEPDHRV**
Band 4	**TKTGYINAAFRSSKNNEAYLFINDKYVLLDYADGTT**
Band 5	**TETQSNNITKFSPDQKNLIFQGDGY**

Figure 1
Extraction and NH_2-terminal sequencing of proteins from *L. culinaris* seeds. Ammonium sulfate saturation fractionation profile of *L. culinaris* seed protein extract: lanes 1–6 correspond to 20, 40, 60, 80 and 90% ammonium sulfate saturation fractions and low molecular-weight markers (Sigma), respectively, on a 15% SDS–PAGE gel stained with Coomassie Blue. The protein bands marked 1–5 from the 80 and 90% fractions were subjected to NH_2-terminal sequencing by Edman degradation and identification of about 20–36 residues was achieved. Band 4 from the 90% ammonium sulfate saturation fraction (lane 6), which corresponds to 2S albumin, was purified and crystallized.

tectant solution. X-ray diffraction intensities were recorded from a single crystal on a MAR345dtb detector (MAR Research) with an incident wavelength of 1.5418 Å generated by a Rigaku RU-H3R Cu rotating-anode X-ray generator equipped with Osmic focusing mirrors. The crystal-to-detector distance, oscillation angle and exposure time were set to 120 mm, 1° and 5 min, respectively. The data images were processed using *AUTOMAR* (Bartels & Klein, 2003) and scaled using *SCALEPACK* (Otwinowski & Minor, 1997).

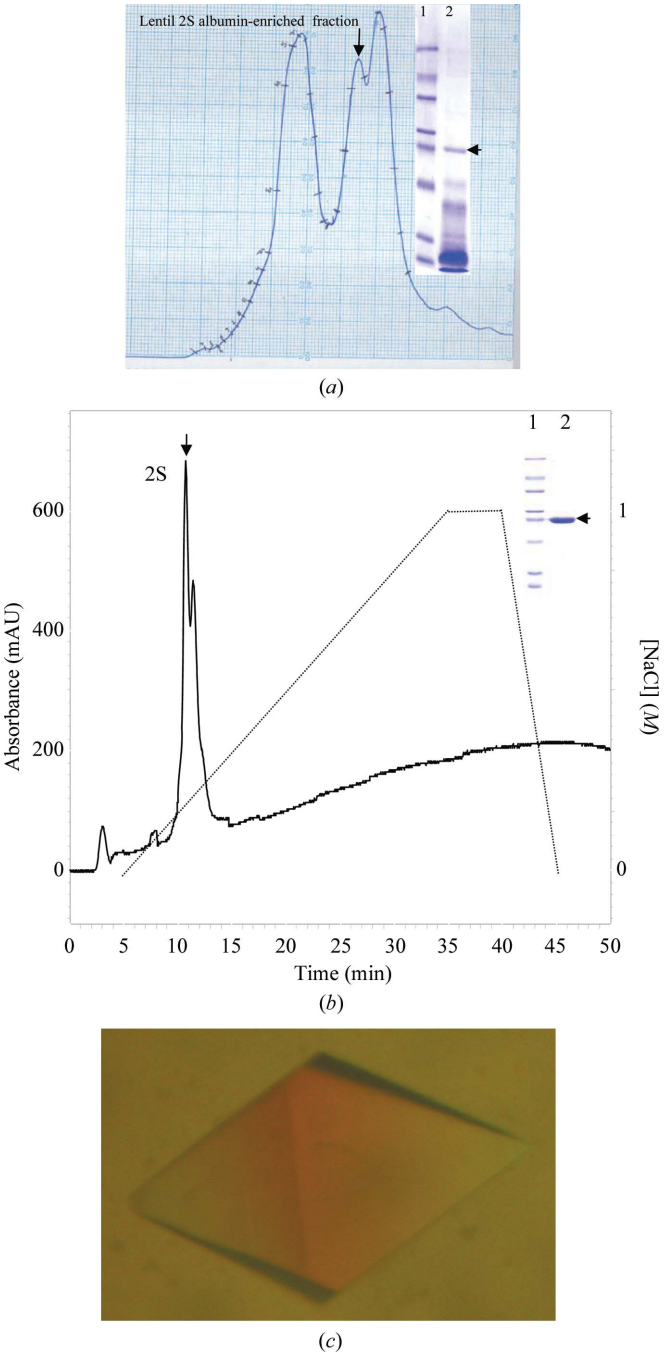


Figure 2
Purification and crystallization of the 2S albumin protein (26.5 kDa) from *L. culinaris* seeds. (*a*) Size-exclusion chromatogram (Sephacryl-300) of the 90% ammonium sulfate saturation fraction. The peak marked with an arrow corresponds to the fraction enriched with the 2S albumin protein from *L. culinaris*. The inset shows a Coomassie Blue-stained 15% SDS–PAGE profile of the protein eluted in the peak marked with an arrow (lane 2) along with low-molecular-weight markers (Sigma; lane 1). (*b*) Weak anion-exchange chromatogram of the 2S albumin-enriched fraction obtained after SEC. The peak corresponding to 2S albumin is marked by an arrow. The dotted line indicates the NaCl gradient. The inset shows a Coomassie Blue-stained 15% SDS–PAGE profile of the purified protein (lane 2) along with low-molecular-weight markers (Sigma; lane 1). (*c*) *L. culinaris* 2S albumin crystal.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P4_1$ (or $P4_3$)
Resolution range (Å)	50–2.5 (2.59–2.50)
No. of observed reflections	89976 (8189)
No. of unique reflections	27943 (2795)
Completeness (%)	98.3 (99.0)
Multiplicity	3.22 (2.93)
$I/\sigma(I)$	5.3 (1.1)
R_{merge} (%)	7.8 (36.7)
Solvent content (%)	
2 molecules per ASU	70.6
3 molecules per ASU	55.9
4 molecules per ASU	40.7

2.5. Mass-spectrometric analysis

Mass-spectrometric analysis was carried out on a VG Platform Electron Spray Ionization mass spectrometer (Fisons Instrument with Quadropole analyzer). 50% acetonitrile was used as mobile phase with nitrogen as the mobile phase carrier. The analyzer was run at a cone voltage of 40 V and a source temperature of 343 K.

3. Results and discussion

NH_2-terminal sequencing led to the identification of various allergy-related proteins from *L. culinaris* (Fig. 1). Four bands designated 1, 2, 3 and 5 were identified from the 80% fraction and a single band (band 4) was identified from the 90% fraction. Bands 1 and 2 (Fig. 1) were identified as convicilin and vicilin, respectively. Within the NH_2-terminal sequence, band 3 showed 79% sequence identity to legumin from *Pisum sativum* and band 5 showed 64% sequence identity to a lectin from *Cicer arietinum* (Fig. 1). Band 4 from the 90% ammonium sulfate saturation fraction was identified as 2S albumin as it showed 85% sequence identity to PA2, a 2S seed albumin from *P. sativum*, and had a molecular weight of 26.522 ± 0.38 kDa as determined by mass-spectrometric analysis.

On purification, 30 mg of pure 2S albumin was obtained from 250 g of *L. culinaris* seeds. Tetragonal bipyramidal shaped protein crystals with typical dimensions of 0.02 × 0.02 × 0.02 mm were obtained from this protein and these crystals diffracted to 2.5 Å resolution (Fig. 2*c*). The crystals belonged to space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 78.6$, $c = 135.2$ Å. The ambiguity in the enantiomorphic space group cannot be resolved on the basis of systematic absences in the diffraction data alone. The data-collection and processing statistics are summarized in Table 1. Briefly, the data scaled to an R_{merge} of 7.85% with an $I/\sigma(I)$ of 5.3 and a completeness of 98.3%. The Matthews coefficient was calculated as 2.1 $Å^3$ Da^{-1}, indicating the presence of four molecules in the asymmetric unit with 40.7% solvent content.

The PA2 seed allergen of *P. sativum* to which band 4 showed homology has been classified as a 2S albumin. The NMR structures of 2S albumins from various sources have been reported, including napin Bn Ib from *Brassica napus* (Rico *et al.*, 1996), Ric C3 from *Ricinus communis* (Pantoja-Uceda *et al.*, 2003), SFA-8 from *Helianthus annuus* (Pantoja-Uceda *et al.*, 2004), Ber e 1 from *Bertholletia excelsa* (Tengel *et al.*, 2005) and Ara h 2 from *Arachis hypogaea* (Lehmann *et al.*, 2006). The 26.5 kDa 2S albumin from *L. culinaris* did

not show any discernible sequence homology to any of the 2S albumins for which structures are known. Since the classification of the protein into the 2S albumin class was on the basis of solubility and sedimentation coefficient (Osborne & Campbell, 1898), the 2S albumin from *L. culinaris* could well be a novel and structurally different protein compared with other known members of the 2S albumin family. Thus, it would be interesting to determine the structure of this protein, which may shed light on the allergenicity of *L. culinaris*. In summary, we have isolated, purified, crystallized and carried out preliminary X-ray characterization of the 2S seed albumin protein from *L. culinaris*.

We thank the Departments of Biotechnology (DBT) and Science and Technology (DST), Government of India for financial support.

References

Bartels, K. S. & Klein, C. (2003). *The AUTOMAR Manual*, v.1.4. Norderstedt, Germany: MAR Research GmbH.
Bischoff, S. & Crowe, S. E. (2005). *Gastroenterology*, **128**, 1089–1113.
Breiteneder, H. & Mills, E. N. (2005). *J. Allergy Clin. Immunol.* **115**, 14–23.
Breiteneder, H. & Radauer, C. (2004). *J. Allergy Clin. Immunol.* **113**, 821–830.
Chanana, V., Kaur, K. J. & Salunke, D. M. (2004). *Acta Cryst.* D**60**, 2100–2103.
Dall'Antonia, Y., Pavkov, T., Fuchs, H., Breiteneder, H. & Keller, W. (2005). *Acta Cryst.* F**61**, 186–188.
Gaur, V., Sethi, D. K. & Salunke, D. M. (2008). *Acta Cryst.* F**64**, 32–35.
Lehmann, K., Schweimer, K., Reese, G., Randow, S., Suhr, M., Becker, W. M., Vieths, S. & Rosch, P. (2006). *Biochem. J.* **395**, 463–472.
López-Torrejón, G., Salcedo, G., Martín-Esteban, M., Díaz-Perales, A., Pascual, C. Y. & Sánchez-Monge, R. (2003). *J. Allergy Clin. Immunol.* **112**, 1208–1215.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, pp. 96–97. New York: John Wiley & Sons.
Madsen, C. (2005). *Proc. Nutr. Soc.* **64**, 413–417.
Osborne, T. B. & Campbell, G. F. (1898). *J. Am. Chem. Soc.* **20**, 348–362.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pantoja-Uceda, D., Bruix, M., Gimenez-Gallego, G., Rico, M. & Santoro, J. (2003). *Biochemistry*, **42**, 13839–13847.
Pantoja-Uceda, D., Shewry, P. R., Bruix, M., Tatham, A. S., Santoro, J. & Rico, M. (2004). *Biochemistry*, **43**, 6976–6986.
Qureshi, I. A., Sethi, D. K. & Salunke, D. M. (2006). *Acta Cryst.* F**62**, 869–872.
Rico, M., Bruix, M. & Rodriguez, R. (1996). *Biochemistry*, **35**, 15672–15682.
Tengel, T., Alcocer, M. J., Schleucher, J. & Larsson, G. (2005). *J. Biomol. NMR*, **32**, 336.