

**Insaf A. Qureshi, Dhruv K. Sethi
and Dinakar M. Salunke***National Institute of Immunology,
New Delhi 110067, India

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

Received 15 May 2006

Accepted 19 July 2006

Purification, identification and preliminary crystallographic studies of an allergenic protein from *Lathyrus sativus*

A 24 kDa protein was purified from the seeds of *Lathyrus sativus* by ammonium sulfate fractionation and ion-exchange chromatography. The N-terminal amino-acid sequence showed significant homology with the 2S albumin class of seed storage proteins. The protein showed 85% sequence homology with the seed albumin of *Pisum sativum* within the 40 N-terminal residues. Crystals were obtained by the hanging-drop vapour-diffusion method. The crystals belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 43.5$, $b = 82.7$, $c = 153.4$ Å.

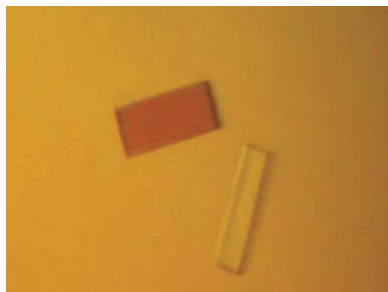
1. Introduction

Allergy is a hypersensitivity reaction initiated by immunological mechanisms, resulting from the interaction between innocuous antigens with antibody and primed T cells produced by an earlier exposure to the same antigen (Johansson *et al.*, 2001). The ability of exogenous proteins to cause respiratory and gastrointestinal allergy and sometimes systemic anaphylactic reactions is well known. Food allergies affect as many as 2–8% of infants and young children, most of whom fortunately outgrow the sensitivity (Helm & Burks, 2000). Although many foods can provoke a reaction, relatively few foods are responsible for the vast majority of food allergic reactions, and include milk, egg, seeds, fish and shellfish (Sampson, 2003). Food allergies are among the most common causes of hypersensitivity, being second only to pollen allergies in number. While pollen allergies have been reasonably well addressed, the biochemical and structural mechanisms underlying food allergies are relatively under-explored.

The comparative structural proteomics of allergens may help establish a structure–function relationship, which in turn would help us understand the physicochemical basis of allergenicity. Many seed storage proteins are known to be responsible for allergies in sensitized individuals (Breiteneder & Ebner, 2001). These allergens induce production of IgE antibodies, leading to mast-cell degranulation and consequent symptomatic responses. Determination of the three-dimensional structures of the allergenic and non-allergenic proteins belonging to common structural families in various plant species would provide insight with regard to the structural basis of allergy.

Grasspea (*Lathyrus sativus*) is an annual pulse crop belonging to the Leguminosae family. Its flour is used mostly as cattle feed and is also occasionally consumed by humans. It may induce symptoms of food allergy and occupational allergy (*e.g.* bronchial asthma, rhinoconjunctivitis, generalized urticaria and facial oedema) in sensitized individuals (Valdivieso *et al.*, 1988; Ortiz *et al.*, 2000; Porcel *et al.*, 2001; Anton Girones *et al.*, 2005). Allergenic proteins with molecular weights of 46, 28 and 21 kDa have been identified by IgE Western blotting in allergic patients (Porcel *et al.*, 2001). However, to the best of our knowledge no allergens from this plant have been fully characterized.

As a part of our continuing efforts to identify and characterize proteins with allergenic properties from legume seeds, this laboratory previously reported the purification and crystallization of a novel



protein from *Vigna unguiculata* (Chanana *et al.*, 2004). Here, we report the identification, purification and crystallization of a 24 kDa protein from *L. sativus*. Sequence analysis indicated that it belongs to the 2S albumin family of seed storage proteins, members of which are known to be intrinsically allergenic.

2. Materials and methods

2.1. Protein purification

Mature seeds of *L. sativus* were purchased from the Indian Agricultural Research Institute (IARI) Regional Station, Karnal, India. The seeds were ground to a fine powder and then defatted with petroleum ether. The defatted powder was homogenized with 10 mM phosphate buffer pH 7.2 containing 140 mM NaCl [1:10(w:v)] by continuous stirring for 4 h at 277 K. The crude extract was prepared by centrifugation at 48 384g for 30 min and then subjected to 0–20, 20–40, 40–60, 60–80 and 80–95% ammonium sulfate fractionation. The pellets obtained after ammonium sulfate fractionation were resuspended in 50 mM Tris–HCl buffer pH 7.5 and dialyzed. All fractions were loaded onto SDS–PAGE gels to identify the proteins. The dialyzed 80–95% fraction was loaded onto a PI/M weak anion-exchange column (Applied Biosystems) to separate the constituent proteins. After loading the sample, the column was initially washed with 50 mM Tris–HCl buffer pH 7.5 for 5 min to separate any unbound protein. The bound protein was subsequently eluted using a gradient of 0–1 M NaCl in 50 mM Tris–HCl buffer pH 7.5 over 35 min. All fractions from ion-exchange columns were precipitated with acetone and checked on a 15% SDS–PAGE gel. The protein concentration was determined by the BCA protein assay (Pierce Biotechnology) using bovine serum albumin as the standard.

2.2. N-terminal amino-acid sequencing

N-terminal amino-acid sequences were obtained by the Edman degradation method on a Procise Protein Sequencer (Applied Biosystems). The protein sequence was subjected to a BLAST database search (<http://www.ncbi.nlm.nih.gov/BLAST>) to find proteins with significant sequence homology.

2.3. Circular-dichroism (CD) analysis

The CD spectrum in the far-ultraviolet region was measured on a Jasco J-710 spectropolarimeter at a scanning speed of 200 nm min⁻¹. A 30 μM sample prepared in 10 mM Tris–HCl buffer pH 7.5 was analyzed at 298 K in a 0.2 cm path-length cuvette.

2.4. Crystallization

For crystallization trials, the protein was used at a concentration of 5 mg ml⁻¹ in 50 mM Tris–HCl buffer pH 7.5. Crystallization was performed at room temperature using the hanging-drop vapour-diffusion method (McPherson, 1982). Crystal Screens I and II (Hampton Research) were used to establish initial crystallization conditions. The drops were composed of equal volumes (4 μl) of protein solution and precipitant solution and were equilibrated against 500 μl reservoir solution. The final crystallization conditions were standardized to be 0.1 M MES pH 6.5 containing 10%(w/v) PEG 20 000.

2.5. Data collection and processing

Data could be collected to a resolution of 3.0 Å from a single crystal using a rotating-anode generator (Rigaku) and a MAR345 dtb detector (MAR Research). Single crystals were soaked in cryo-

protectant solution [33%(v/v) glycerol, 10% PEG 20 000, 0.1 M MES pH 6.5] for 60 s and mounted in nylon cryoloops. The pre-soaked crystals were immediately subjected to flash-freezing in a cold nitrogen stream. The data were collected at 120 K with a crystal-to-detector distance of 120 mm, an oscillation range of 1° and an exposure time of 1 min per image.

3. Results and discussion

Crude extract and ammonium sulfate fractions were analyzed on 15% SDS–PAGE (Fig. 1*a*). N-terminal amino-acid sequencing of several electrophoresed proteins was carried out subsequent to transfer onto a PVDF membrane. N-terminal sequencing led to the identification of many allergy-related proteins. The N-terminal amino-acid sequences of bands 1, 2 and 4 from the 80–95% ammonium sulfate fraction had homology with the 7S vicilin storage proteins of *Pisum sativum* and *Vicia faba*, while band 3 showed significant homology with other plant albumins (Fig. 1*b*) that are known to be associated with allergies (Porcel *et al.*, 2001).

When the resuspended pellet from the 80–95% ammonium sulfate fractionation was loaded onto a PI/M weak anion-exchange column, a protein was eluted in the flowthrough. The other proteins from the 80–95% fraction were eluted at various concentrations of salt in a 0–1 M NaCl gradient (Fig. 2*a*). The eluted protein in the flowthrough appeared as a single band with a molecular weight of 24 kDa and corresponded to band 3 when analyzed on a 15% SDS–PAGE gel (Fig. 2*b*). While attempts are being made to sequence the entire protein by obtaining proteolytic fragments, a BLAST search of the N-terminal 40-amino-acid sequence of the purified protein against the protein sequences in the PDB was carried out (Fig. 2*c*). The search revealed that this protein shows about 85% sequence homology with the seed albumin of *P. sativum*. Based on the sequence homology, the 24 kDa protein was classified as belonging to

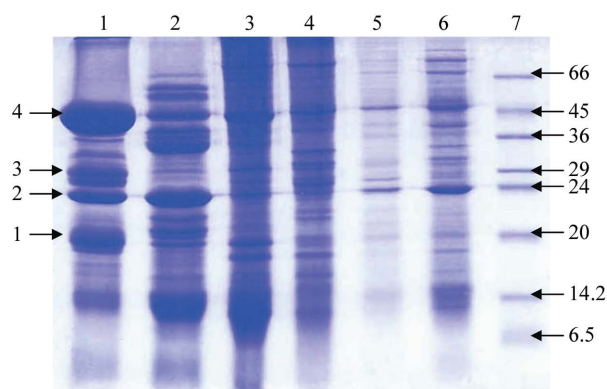


Figure 1
 Extraction and N-terminal sequences of proteins from seeds of *L. sativus*. (*a*) Ammonium sulfate fractionation of protein extract. Lanes 1–7 correspond to 95, 80, 60, 40, 20% ammonium sulfate, crude and low molecular-weight markers (Sigma), respectively. (*b*) N-terminal sequences of the four dominant proteins from the 80–95% fractions.

Band 1	TSRQENENPFIRESNQFQTLNEN
Band 2	DKRQEINEEAIVVKKLREQQ
Band 3	TKPGYINAAFRSSKNNEAYFFINDK YVLLDYAPGSXRDKV
Band 4	RSDQENPFIFKSNFFQTLYENEN

the albumin family of seed storage proteins. The 2S albumins are a major group of storage proteins present in many dicotyledonous plant species. They are low-molecular-weight proteins that are generally composed of two different polypeptide chains linked by two disulfide bonds (Shewry *et al.*, 1995). Several tree nut and seed allergens are 2S albumins. They include Ber e 1 from brazil nut, Jug r 1 from English walnut, Ana o 3 from cashew nut, Ses i 2 from sesame seeds, Sin a 1 from yellow mustard seeds and Bra j 1 from oriental mustard seeds (Breiteneder & Radauer, 2004).

The secondary structure of the purified protein was investigated by circular-dichroism analysis. The spectrum is shown in Fig. 2(d) and has a minimum at 218 nm with a mean residue molar ellipticity of $-3357 \text{ deg cm}^2 \text{ dmol}^{-1}$. It can be inferred that the protein predominantly contains β -sheet structure.

Initially small crystals were obtained with polyethylene glycols of molecular weight ranging from 3350 to 20 000 Da. However, these crystals were not suitable for X-ray diffraction analysis. To optimize

the crystallization conditions, changes in precipitant and pH range were explored. The best crystals grew in 0.1 M MES pH 6.5 containing 10% (w/v) PEG 20 000 within 10 d (Fig. 2e) to maximum dimensions of approximately $0.25 \times 0.1 \times 0.05 \text{ mm}$ and diffracted to 3.0 Å resolution. The crystals belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 43.5$, $b = 82.8$, $c = 153.4 \text{ Å}$. Data in the 50.00–3.00 Å resolution range were processed and analyzed using *AUTOMAR*. The data-collection and processing statistics are summarized in Table 1. Briefly the data scaled to an R_{merge} of 12.7 (36.3) with an $I/\sigma(I)$ of 3.6 (1.4) and a completeness of 97.6%. Assuming the presence of two monomers in the asymmetric unit, the calculated Matthews coefficient was $2.87 \text{ Å}^3 \text{ Da}^{-1}$ and the solvent content was 57%.

The three-dimensional structure of 2S albumin from seeds, *e.g.* napin Bn Ib from rape (Rico *et al.*, 1996), Ric C3 from castor bean (Pantoja-Uceda *et al.*, 2003) and SFA-8 from sunflower (Pantoja-Uceda *et al.*, 2004), have been determined by NMR methods. These

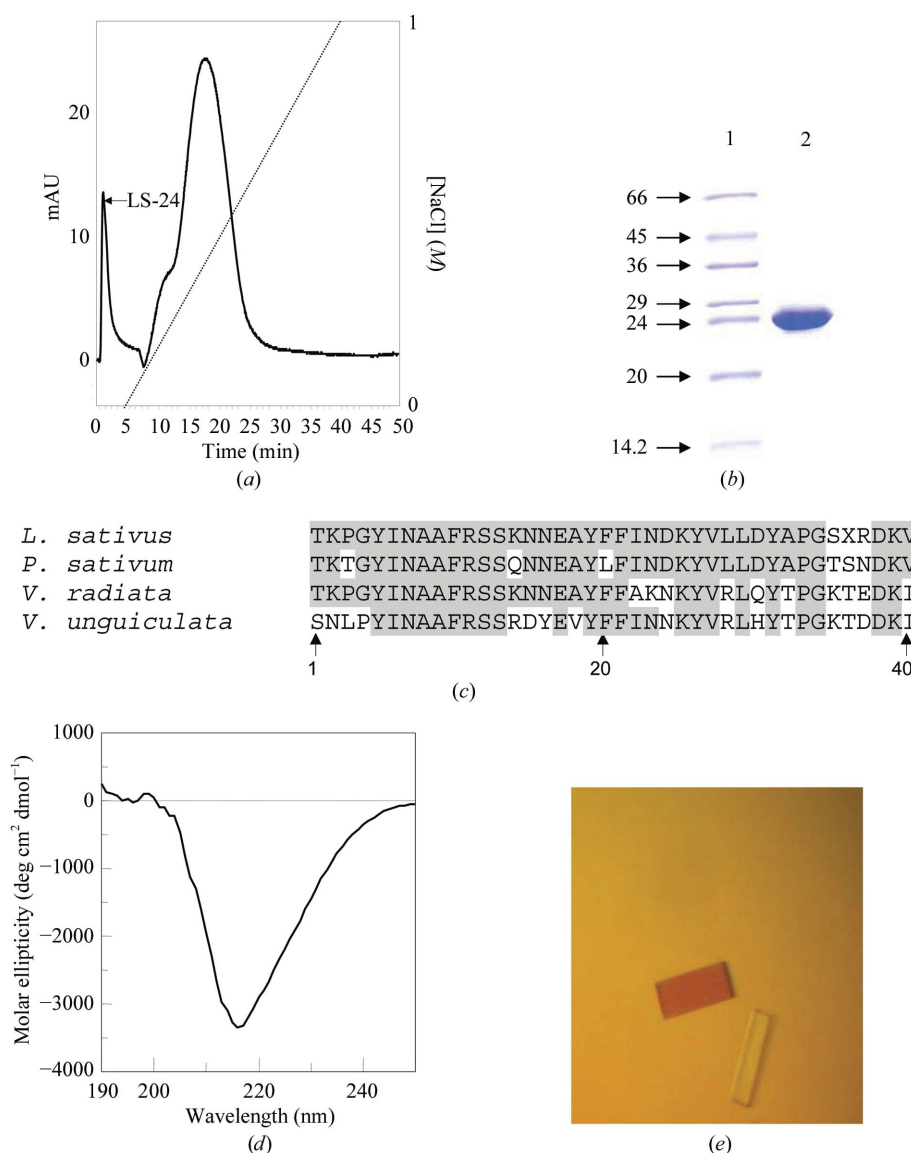


Figure 2

Purification, characterization and crystallization of 24 kDa protein from seeds of *L. sativus*. (a) Purification of 24 kDa protein by ion-exchange chromatography on a PI/M column. The peak LS-24 corresponds to the 24 kDa protein from *L. sativus*. The dotted line indicates the NaCl gradient. (b) Purified 24 kDa protein on SDS-PAGE. Lanes 1 and 2 correspond to protein markers (kDa) and purified protein, respectively. (c) Sequence alignments of the 24 kDa protein of *L. sativus* with other legume albumins. Identical amino-acid residues are highlighted in dark grey. (d) CD spectra of the 24 kDa protein in the far-UV region. (e) Crystals of the 24 kDa protein grown by the hanging-drop method.

Table 1

Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_1$
Unit-cell parameters (Å)	$a = 43.5, b = 82.8, c = 153.4$
Temperature (K)	120
Wavelength (Å)	1.5418
Oscillation steps	101
Crystal-to-detector distance (mm)	120
Resolution limit (Å)	50.0–3.0
Exposure time per image (s)	60
No. of observed reflections	46241
Total No. of unique reflections collected	11436
Completeness (%)	97.6 (94.3)
R_{merge}	12.7 (36.3)
No. of molecules in ASU	2
$\langle I/\sigma(I) \rangle$	3.6 (1.4)
Matthews coefficient (Å ³ Da ⁻¹)	2.87
Solvent content (%)	57

proteins did not show any significant sequence homology with the 24 kDa protein of *L. sativus*.

No structures of the proteins with which the 24 kDa protein of *L. sativus* shows homology have been determined so far. Fig. 2(c) shows that the 40 N-terminal amino-acid residues of the 24 kDa protein of *L. sativus* have significant homology with corresponding proteins from *P. sativum* (85%) and *V. unguiculata* (66%). On the other hand, the amino-acid sequence showed no homology with the sequences of other 2S albumins such as napin, Ric C3 and SFA-8. Additionally, these proteins are substantially smaller in size (12–15 kDa). Therefore, the 24 kDa protein from *L. sativus* is likely to be significantly different from napin, Ric C3 or SFA-8. The structure determination of the 24 kDa protein of *L. sativus* may shed light on the structural basis of the allergenicity of *L. sativus* flour.

In summary, we have isolated, purified, crystallized and carried out preliminary X-ray characterization of the 24 kDa allergenic protein

from *L. sativus*. The structure solution and refinement of this protein, which are in progress, will provide important insights into the possible structural role in its allergenic properties.

We thank the Department of Biotechnology, Government of India for financial support.

References

- Anton Girones, M., de la Hoz Caballer, B., Munoz Martin, T., Cuevas Agustin, M. & Sanchez-Cano, M. (2005). *Allergol. Immunopathol. (Madr.)*, **33**, 326–328.
- Breiteneder, H. & Ebner, C. (2001). *Curr. Opin. Allergy Clin. Immunol.* **1**, 261–267.
- Breiteneder, H. & Radauer, C. (2004). *J. Allergy Clin. Immunol.* **113**, 821–830.
- Chanana, V., Kaur, K. J. & Salunke, D. M. (2004). *Acta Cryst.* **D60**, 2100–2103.
- Helm, R. M. & Burks, A. W. (2000). *Curr. Opin. Immunol.* **12**, 647–653.
- Johansson, S. G., Hourihane, J. O., Bousquet, J., Brujinzeel-Koomen, C., Dreborg, S., Haahtela, T., Kowalski, M. L., Mygind, N., Ring, J., van Cauwenberge, P., van Hage-Hamsten, M. & Wüthrich, B. (2001). *Allergy*, **56**, 813–824.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: Wiley.
- Ortiz, G. G., Conde-Salazar, L., Gimaraens, D., de la Hoz, C. & Agustin, M. C. (2000). *Contact Derm.* **43**, 43.
- 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.
- Porcel, S., Leon, F., Valero, A. M., Martín-Calderin, P., Cuevas, M. & Alvarez-Cuesta, E. (2001). *J. Allergy Clin. Immunol.* **107**, 743–744.
- Rico, M., Bruix, M. & Rodriguez, R. (1996). *Biochemistry*, **35**, 15672–15682.
- Sampson, H. A. (2003). *J. Allergy Clin. Immunol.* **111**, S540–S541.
- Shewry, P. R., Napier, J. A. & Tatham, A. S. (1995). *Plant Cell*, **7**, 945–956.
- Valdivieso, R., Quirce, S. & Sainz, T. (1988). *Allergy*, **43**, 536–539.