Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, identification and preliminary crystallographic characterization of a novel seed protein from *Vigna unguiculata*

Veenu Chanana, Kanwal J. Kaur and Dinakar M. Salunke*

National Institute Of Immunology, Aruna Asaf Ali Marg, New Delhi, India

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

Received 8 August 2004
Accepted 23 September 2004

A tropical legume, *Vigna unguiculata*, was explored in order to identify potential allergens among the abundant seed proteins and to attempt their crystallographic study. Salt fractionation of the seed extract followed by chromatographic separation led to the purification of a 25 kDa protein. Gel-filtration chromatography of the 80% ammonium sulfate precipitation fraction led to separation of this protein in pure form, which was subjected to N-terminal sequencing. The N-terminal sequences of internal fragments of this protein showed 85% homology to mung bean seed albumin. This family of proteins is known to be intrinsically allergenic. Rhombic shaped crystals were obtained that diffracted to about 2.1 Å resolution. The crystals belong to space group *C*2 and have unit-cell parameters $a = 124.9$, $b = 60.1$, $c = 67.5$ Å, $\beta = 111.1^\circ$.

1. Introduction

The abundantly utilized legume seeds are a common source of allergy in about 4–10% of the population of the world. Atopic allergy is a genetically determined hypersensitivity that is based on the production of IgE antibodies against essentially innocuous antigens (Onaderra *et al.*, 1994). The immediate symptoms (*e.g.* allergic rhinitis, conjunctivitis, asthma, dermatitis, anaphylactic shock) result from allergen-induced cross-linking of effector-cell-bound IgE antibodies and subsequent release of biologic mediators (*e.g.* histamine, leukotriene; Heyek *et al.*, 1998; Segal *et al.*, 1997). Many legumes have been identified to cause allergenic diseases on ingestion (Monsalve *et al.*, 1993). A number of characteristics render a food capable of provoking systemic allergic reaction, such as its ability to stimulate high titres of IgE and to resist gastrointestinal degradation sufficiently to produce fragments containing multiple IgE-binding epitopes (Maleiki *et al.*, 2000). Thus, biochemical and structural characterization could facilitate the understanding of the disease processes associated with the allergens (Teuber *et al.*, 1998). Allergens are usually resistant to proteases, heat and denaturants, allowing them to resist degradation during food preparation and digestion.

Our approach has been to identify proteins with allergenic potential from different commonly used seed sources by correlating their sequences with those of known allergens and subjecting the relevant proteins to crystallographic studies. We identified such a protein in the seeds of the tropical legume *Vigna unguiculata* (cowpea). We extracted proteins from these seeds and selected abundantly present proteins for sequencing and purification. A 25 kDa protein was identified on the basis of N-terminal sequencing of four internal fragments obtained by tryptic digestion. The protein belongs to the 2S albumin family of storage proteins in seeds. This family of proteins has been intrinsically linked to allergy-causing potential (Pastorello *et al.*, 2001).

2. Experimental methods

2.1. Protein purification

The isolation of proteins from 100 g of *V. unguiculata* seeds was carried out using 20 m*M* Tris–HCl buffer pH 7.5 containing 150 m*M* NaCl as homogenization buffer. The crude extract obtained was then centrifuged at 16 900*g* for 40 min at 277 K and subjected to salt fractionation over a concentration range of 20–80%(*w*/*v*) ammonium sulfate, separating the proteins based on their differential solubilities. The pellets obtained were resuspended in 50 m*M* Tris–HCl buffer pH 7.5 and dialyzed to remove salt. SDS–PAGE analysis to identify the proteins was performed according to the Mini Protein II dual-gel method (Bio-Rad). The isolation and purification of proteins were carried out at 277 K using a combination of chromatographic techniques including gel-filtration (Sephacryl S-200, Amersham Pharmacia Biotech) and ion-exchange (PI/M weak anion exchanger, Pharmacia) chromatography. The purified protein concentration was determined by the Bio-Rad protein assay.

2.2. Protein-sequence analysis

Amino-acid sequence determination was performed by the Edman degradation method on a Procise (Applied Biosystems), Protein Sequencer. The protein-sequence homology with the internal peptide sequences obtained on tryptic digestion was determined using a BLAST database search (http://www.ncbi.nlm.nih.gov/BLAST/).

2.3. Crystallization

The purified protein was used in 50 m*M* Tris–HCl buffer pH 7.5. The protein was adjusted to a concentration of 10 mg ml^{-1} for crystallization and for several other biochemical characterizations. The protein was centrifuged to remove the insoluble components. Crystallization was performed at room temperature using conventional hanging-drop vapour diffusion (McPherson, 1982). Initial crystallization experiments were set using Crystal Screen HR2-110 (supplied by Hampton Research). The protein gave crystals with most reagents from the screen that contained PEG 3K or PEG 8K in the range 14%–30%(*w*/*v*). The crystallization conditions were therefore optimized with PEG 3K or PEG 8K as precipitants.

2.4. X-ray diffraction analysis

A protein crystal was soaked for about 10 s in cryoprotectant solution [10%(*v*/*v*) glycerol, 30%(*w*/*v*) PEG 8K, 50 m*M* Tris–HCl] and then transferred to the goniometer under a cryo-jet (Oxford) for data collection. X-ray diffraction intensities were then recorded on an image-plate detector (MAR Research, Germany) with an incident X-ray wavelength of 1.5418 Å generated by a Cu source on a Rigaku rotating-anode generator at 120 K. The crystal-to-detector distance, oscillation angle and exposure time for each image frame were set to 120 mm, 1° and 5 min, respectively.

2.5. Mass-spectroscopic analysis

The mass-spectroscopic analysis was carried out on a VG Platform Electron Spray Ionization mass spectrometer (Fisons instrument with Quadropole analyzer). The mobile phase used was 50% acetonitrile 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.)

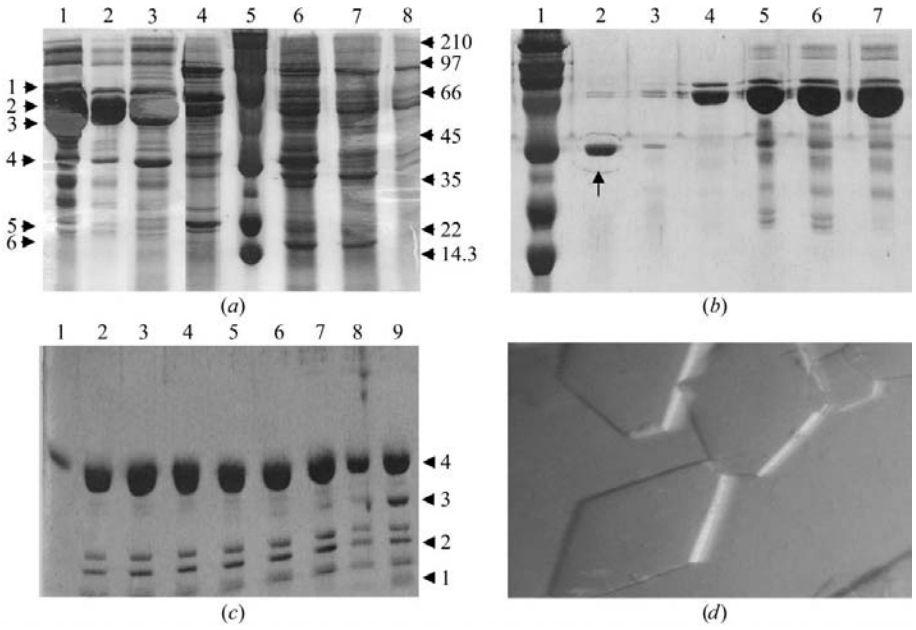

Figure 1
Purification, characterization and crystallization of the 25 kDa protein from *V. unguiculata* seeds. (*a*) Differential protein precipitation by ammonium sulfate salt fractionation. Lanes 1–8 correspond to crude, 90, 80, 70% salt fractions, RPN756 molecular-weight markers (in kDa on right; Sigma), 60, 40 and 20% salt fractions, respectively. (*b*) S-200 column-chromatographic purification profile on SDS–PAGE of the 25 kDa protein using the 80% ammonium sulfate fraction. The protein was eluted in 50 m*M* Tris–HCl buffer. Lanes 1–7 correspond to RPN756 molecular-weight markers and fraction numbers 6, 5, 4, 3, 2 and 1, respectively. (*c*) SDS–PAGE analysis of trypsin digestion of the N-terminal blocked 25 kDa protein for 11 h. Samples were taken each hour and the predominant bands of digested proteins were used for N-terminal sequence identification. Lanes 1–9 correspond to crude, 0, 1, 2, 3, 4, 5, 7 and 11 h fractions, respectively. (*d*) The crystals of the 25 kDa protein used for diffraction studies.

Figure 2
The sequence identification and homology-based characterization of the 25 kDa N-terminal blocked protein. (*a*) The N-terminal sequences for four dominant protein fragments obtained by trypsin digestion. (*b*) The sequence alignments of the four protein fragments to a common protein sequence belonging to the albumin family from *V. radiata*. Residues that are non-identical between the two proteins are marked in red.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the last resolution shell (2.18–2.1 Å).

Space group	*C*2
Unit-cell parameters (Å, °)	a = 124.9, b = 60.1, c = 67.5, β = 111.1
Temperature (K)	120
Wavelength (Å)	1.5418
Oscillation step (Å)	1
Crystal-to-detector distance (mm)	120
Resolution limit (Å)	100–2.1
Exposure time per image (s)	300
Total observations	112436 (2482)
Unique reflections	27508 (490)
Completeness (%)	97.4 (91.2)
$R_{\rm merge}$† (%)	10.5 (32.2)
Protomers per AU	2
$V_{\rm M}$ (Å^3 Da^{-1})	2.33
Solvent content (%)	47.11

† $R_{\rm merge} = \sum_{hkl}\sum_j I(hkl)_j - \langle I(hkl)\rangle / \sum_{hkl}\sum_j I(hkl)_j$, where $I(hkl)_j$ is the jth measurement of the intensity of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl.

3. Results

3.1. Biochemical characterization

The protein fractions obtained on *V. unguiculata* seed extraction were analyzed on 12% SDS–PAGE (Fig. 1*a*). A range of proteins with significant molecular-weight differences (as indicated by arrows in Fig. 1*a*) were identified in various protein fractions obtained by adding salt. These proteins transferred onto a PVDF membrane were individually subjected to N-terminal sequencing, which helped to identify the homologues, if any, from database searches using N-terminal segments, effectively leading to the characterization of these proteins. Several allergy-related proteins were identified, purification of which for subsequent crystallographic studies is in progress. The protein sample obtained with 80% salt fraction was loaded onto a Sephacryl S-200 column (Amersham Pharmacia Biotech) pre-equilibrated with 50 m*M* Tris–HCl buffer pH 7.5. The purified protein corresponds to a molecular weight of 25 kDa when analyzed on 12% SDS–PAGE using molecular-weight markers. The 25 kDa protein appeared as a single band on chromatographic separation of the 80% salt fraction when loaded onto a Sephacryl S-200 column (Fig. 1*b*). The same protein was eluted in a pure form as an unbound protein fraction when loaded onto a PI/Ma anion-exchange column (Pharmacia).

The chromatographically eluted fractions were collected and prepared for subsequent study with a final concentration of 10 mg ml^{-1} (determined by the Bio-Rad protein assay with BSA as the standard marker). The protein was concentrated using a 10 kDa Centriplus (Millipore). The mass-spectrometric analysis suggested that albumin has a molecular weight of 25.409 ± 0.009 kDa.

3.2. Sequence-based analysis

The N-terminal sequencing of the 25 kDa protein suggested the possibility of the N-terminus being blocked, although purification was relatively easy. Therefore, in order to identify the protein, internal peptide fragments were generated by digestion of the pure protein with trypsin at 310 K for 15 h (enzyme:protein at a ratio of 1:70) in 100 m*M* ammonium bicarbonate buffer pH 8.5. The aliquots were taken at every 1 h interval and the reaction was stopped by adding reducing dye (Coomassie brilliant blue R-250) and boiling the sample. The aliquots taken at each interval were then analyzed on SDS–PAGE (Fig. 1*c*) and subsequently sequenced. The protein-sequence homology was determined using the N-terminal sequences of the internal peptide fragments obtained using a BLAST database search (http://www.ncbi.nlm.nih.gov/BLAST/).

Fig. 2(*a*) shows the N-terminal sequences of the four internal peptide fragments. All the four peptides could be aligned with a single protein, plant albumin from *V. radiata* seeds (Fig. 2*b*). Within the known sequence regions, the 25 kDa query sequence showed about 85% sequence identity with the plant albumin from *V. radiata*. It was therefore inferred that the 25 kDa protein belonged to the albumin family of seed storage proteins.

3.3. Crystallization and characterization of the protein crystals

A large number of attempts were made to obtain crystals of the protein under different pH and ionic strength conditions using a variety of different precipitants and protein concentrations. Successful crystallization was achieved when 5 µl protein solution was mixed with 5 µl precipitant solution (50 m*M* Tris–HCl pH 7.5 containing 10–30% PEG 8K) and equilibrated against the precipitant solution. Rhombic shaped 25 kDa protein crystals were obtained with 30% PEG 8K (Fig. 1*d*).

Diffraction data were collected to a 97.4% completeness at 2.1 Å resolution. The data were processed using *DENZO* and combined using *SCALEPACK* (Otwinowski & Minor, 1997). Systematic analysis suggested that the crystals belong to the monoclinic space group *C*2, with unit-cell parameters a = 124.8, b = 60.1, c = 67.5 Å, β = 111.1°. The details of the data collected and the processing statistics are given in Table 1. Assuming the presence of two protomers in the asymmetric unit, the Matthews coefficient ($V_{\rm M}$) was calculated to be 2.33 Å^3 Da^{-1} and the solvent content was calculated to be 47.11%.

4. Discussion

From our inital aim to crystallize seed allergens, we have succeeded in purifying, identifying and crystallizing a 25 kDa protein from *V. unguiculata* seeds associated with allergenicity. The purified 25 kDa protein was N-terminally protected; therefore, it was characterized by sequencing the internal peptide fragments generated by trypsin digestion. Characterization of the protein based on the sequence correlations using the four internal fragments helped to identify the protein as belonging to the albumin family of seed storage proteins. Plant albumins are a family of low-molecular-weight (15–20 kDa) seed storage proteins (Shewry *et al.*, 1995). These proteins are abundantly synthesized as a single precursor protein that is cleaved post-translationally into a molecule composed of two different polypeptide chains linked to two disulfide bridges. The albumins were initially classified on the basis of their sedimentation coefficient (S20) as proteins with an S20 of ∼2 (Osborne & Campbell, 1898). These proteins are characterized by their solubility at their isoelectric points in the absence of salt. These proteins show great variations in their subunit structure and synthesis; however, they are all compact globular proteins with conserved cysteine residues spaced in the sequence ...C... C.../...CC...C...C*X*C...C...C that are involved in making this protein highly stable and compact (Pantoja-Uceda *et al.*, 2002).

Only two 2S albumins, napin (Bnlb) from rape seed (Rico *et al.*, 1996) and Ric C3 from castor bean seeds, have been structurally characterized by NMR (Pantoja-Uceda *et al.*, 2003). These proteins show no homology with the 25 kDa albumin. Neither the structure of the albumin from *V. radiata* nor that of any of the proteins that have high sequence homology to it have been reported so far. Therefore, our determination of the structure of the 25 kDa albumin family protein from *V. unguiculata* may provide a novel protein structure. Our crystals diffract to a resolution of 2.1 Å and are stable in the X-ray beam. As no crystal structures of closely related proteins are available, structure determination is being pursued following *ab initio* phase determination using heavy-atom derivatives.

Generally, the 2S albumins have been observed to be highly compact, which possibly confers stability against thermal denaturation and digestion by proteolytic enzymes on them. Sin a I, a major food allergen from mustard seeds, has been shown to exhibit high stability even at low pH values, together with an ability to interact with membranes (Onaderra *et al.*, 1994). There may be an underlying structural property associated with the allergenic character of the 2S albumins, since resistance to digestion and interaction with membranes are the key factors for allergenicity of food components. Determination of the structure of the 25 kDa protein from *V. unguiculata* will provide further insights in correlating structure and allergenic behaviour.

References

Heyek, B., Vangelista, L. & Paslore, A. (1998). *J. Immunol.* **161**, 7031–7039.

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: Wiley.

Maleiki, S. J., Kopper, R. A., Sheer, D. S. & Bannon, G. A. (2000). *J. Immunol.* **164**, 5884–5889.

Monsalve, R. I., Manuel, A. & Rodriguez, R. (1993). *Biochem. J.* **293**, 625–632.

Onaderra, M., Monsalve, R. I. & Rodriguez, R. (1994). *Eur. J. Biochem.* **225**, 609–615.

Osborne, T. B. & Campbell, G. F. (1898). *J. Am. Chem. Soc.* **20**, 348–405.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pantoja-Uceda, D., Bruix, M., Rico, M. & Santoro, J. (2003). *Biochemistry*, **42**, 13839–13847.

Pantoja-Uceda, D., Bruix, M., Santoro, J., Rico, M. & Villalba, M. (2002). *Biochem. Soc. Trans.* **30**, 919–924.

Pastorello, E. A., Pompei, C., Pravettoni, V., Brenna, O., Farioli, L., Trambaioli, C. & Conti, A. (2001). *Allergy*, **5**, 45–47.

Rico, M., Bruix, M. & Rodriguez, R. (1996). *Biochemistry*, **35**, 15672–15682.

Segal, D. M., Taurog, J. D. & Metzger, H. (1997). *Proc. Natl Acad. Sci. USA*, **74**, 2993–2997.

Shewry, P. R., Napier, J. A. & Tatham, A. S. (1995). *Plant Cell*, **7**, 945–956.

Teuber, S. S., Dandekar, A. M., Peterson, W. R. & Sellers, C. L. (1998). *J. Allergy Clin. Immunol.* **101**, 807–814.