

Solution structure of allergenic 2 S albumins

D. Pantoja-Uceda*, M. Bruix*, J. Santoro*, M. Rico*[†], R. Monsalve[‡] and M. Villalba[‡]

*Instituto de Química Física Rocasolano, CSIC, Serrano 119, Madrid 28006, Spain, [†]ALK-Abelló, Miguel Fleta 19, Madrid 28037, Spain, and [‡]Departamento de Bioquímica y Biología Molecular, Universidad Complutense de Madrid, Madrid 28040, Spain

Abstract

The NMR solution structures at different levels of refinement of three different 2 S albumin seed proteins, the recombinant pronapin precursor from *Brassica napus*, the recombinant RicC3 from *Ricinus communis* and the methionine-rich protein from sunflower (*Helianthus annuus*), are described. The resulting common structure consists of a bundle of five α -helices, folded in a right-handed superhelix. The structure is very similar to that of other plant proteins: the hydrophobic protein from soybean, non-specific lipid transfer proteins and amylase/trypsin inhibitors. Analogies and differences in the structures of these families, as well as their possible relationship to allergenicity, are discussed.

Introduction

The 2 S albumins are storage proteins that are widely distributed in plant seeds. They are small proteins (12–15 kDa) that are generally composed of two different polypeptide chains linked by two disulphide bridges. The most characteristic feature of their amino acid sequence is the distribution of their eight cysteine residues in a conserved pattern (...C...C.../...CC...CXC...C...C). In addition to the two interchain disulphide bridges, there are two other intrachain disulphide bonds, which make these proteins very stable and compact. Due to their amino acid composition, their high content in the protein bodies of seeds and their mobilization during germination, a role as a nitrogen and sulphur donor has been proposed for these proteins [1]. However, other activities have been

ascribed to the 2 S albumins: they have been shown to act as antifungal agents [2], serine protease inhibitors and calmodulin antagonists [3]. In addition to their biochemical interest, the 2 S albumins have been used by means of genetic engineering as carriers for the synthesis of biologically active peptides [4], as well for improving the nutritional properties of grain crops by increasing their content of essential amino acids [5]. On the other hand, one of the products of the 2 S seed storage protein from *Ricinus communis*, named RicC3 [6], constitutes the peptidic component of the immunomodulator Inmunoforon[®], a widely used pharmaceutical speciality [7].

Apart from these actual or potential biochemical properties of 2 S albumins, what is most interesting in the framework of this colloquium is their allergenic properties. 2 S albumins from different species, i.e. those from yellow and oriental mustard, rapeseed and others [8–11], have been known for many years to be allergenic. Moreover, in recent years, 2 S albumins from several other species have been described as being allergens [12], suggesting that this family of storage proteins is intrinsically allergenic.

To date, there has been only one structural study of a 2 S albumin, that of napin BnIb [13], for which the residue heterogeneity present in the wild-type protein isolated from rapeseed precluded the determination of a high-resolution three-dimensional structure, and only the global fold could be determined. Recently, the high-yield synthesis by recombinant methods of a napin from rapeseed [14] and of RicC3 from castor bean [15] has opened the way to a more precise determination of their three-dimensional structures, which may provide insights into analogies and differences between the behaviour of albumins from different species. Moreover, a joint study of the three-dimensional structure of the several allergenic 2 S albumins together with those of the structurally related lipid transfer proteins (LTPs), which are also allergenic, would be a great help in establishing meaningful relationships between structure and allergenic behaviour. Progress in the

Key words: *Brassica napus*, *Helianthus annuus*, NMR, *Ricinus communis*, seed proteins.

Abbreviations used: HPS, hydrophobic protein from soybean; HSQC, heteronuclear single-quantum coherence; LTP, lipid transfer protein; nsLTP, non-specific lipid transfer protein; NOE, nuclear Overhauser effect; SFA-8, methionine-rich 2 S albumin 8 from sunflower.

[†]To whom correspondence should be addressed (e-mail mrco@iqfr.csic.es).

structural determination of some 2 S albumins will be reported here.

Materials and methods

We have studied the structural features of three different 2 S albumin seed proteins: the recombinant pronapin precursor from *Brassica napus* (proBnIb), recombinant RicC3 from *Ricinus communis*, and the methionine-rich 2 S albumin 8 from sunflower seeds (*Helianthus annuus*), SFA-8.

Expression of the cDNA encoding the precursor form of napin BnIb (proBnIb) in the yeast *Pichia pastoris* provided, after purification by size-exclusion chromatography and reverse-phase HPLC, a protein including 31 residues of the small chain and 75 of the large chain, connected by the segment Ser-Glu-Asn [14]. Several milligrams of the protein were given to us as a gift by the authors of [14].

Recombinant ^{15}N -labelled RicC3 from *R. communis* has been produced by using a new system for high-level expression of heterologous proteins in native conformation using minimal medium cultures of *Escherichia coli* [15] and $^{15}\text{NH}_4\text{Cl}$ as nitrogen source. This ^{15}N -labelled sample, provided to us by the authors of [15], facilitated the process of resonance and NOE (nuclear Overhauser effect) assignments by using three-dimensional NMR methods. Recombinant Ric3C is a 125-residue protein containing the short (residues 8–39) and long (residues 56–120) chains of the mature protein, together with a 16-residue linker (40–55) between the two chains as well as extensions at the N- and C-termini of seven and five residues respectively.

Several milligrams of sunflower albumin SFA-8, purified from sunflower seeds and prepared as described in [16], were given to us by the authors of [16]. Mature SFA-8 is a single-chain protein of 103 amino acids, containing an unusually large proportion of hydrophobic residues, including 16 methionines.

Samples were prepared for NMR experiments at 1–2 mM concentration in 95 % H_2O /5 % $^2\text{H}_2\text{O}$ or in $^2\text{H}_2\text{O}$ solution containing 10 mM sodium phosphate and (trimethylsilyl)propionic acid (internal reference) at pH 3.5 in 5 mm NMR tubes. The experiments performed included two-dimensional ^1H NOESY (80 ms mixing time), ^1H TOCSY (65 ms mixing time), ^1H COSY and, in the case of RicC3, two-dimensional $^1\text{H}/^{15}\text{N}$ hetero-nuclear single-quantum coherence (HSQC), and three-dimensional ^{15}N NOESY-HSQC (80 ms

mixing time), ^{15}N TOCSY-HSQC (65 ms mixing time), and experiments that correlate the intraresidue ^1HN and $^1\text{H}\alpha$ resonances and the intraresidue ^{15}N and $^1\text{H}\beta$ resonances. All NMR data were collected at 308 K on a Bruker DMX 600 NMR spectrometer.

Results

The assignment of resonances of proBnIb, carried out by a combination of two-dimensional experiments following a standard strategy, is now almost complete. About 600 non-trivial NOE cross-correlations have been also assigned, which are being used for determining a more precise three-dimensional structure than the one we obtained some years ago on the basis of a protein from natural sources. In the previous study [13], low-intensity satellite peaks were found for quite a number of TOCSY cross-correlations involving residues in well-defined segments of the protein. This was ascribed to residue heterogeneity, most probably located at residues 32 and 48 of the long chain, which was the main cause of ambiguities in the process of NOE assignment. This low-intensity second set of signals is missing from the spectrum of the recombinant protein, thus confirming that it arises from polymorphisms present in the natural protein. Notwithstanding these differences, the resonance assignment of the recombinant protein matches that described previously very closely [13], validating the conclusions drawn in our previous report.

The resonance assignment of ^1H and ^{15}N resonances of RicC3 was achieved by a combination of two- and three-dimensional experiments using standard methods [17]. The resulting assignment of ^1H and ^{15}N resonances has been deposited in the BioMagResBank with accession number BMRB-5374. The three-dimensional solution structure of RicC3 was determined on the basis of ~ 1550 distance constraints derived from unambiguously assigned NOE cross-correlations supplemented with 70 ϕ and 55 ψ torsion angle constraints arising from homo- and hetero-nuclear vicinal couplings. The average pairwise root-mean-square distance for the backbone heavy atoms, excluding residues in the random coil, is 0.92 Å. A full account of the NMR structure determination will be given elsewhere. The resulting three-dimensional structure (see Figure 1c) consists of five helices spanning residues 18–23, 30–35, 63–72, 78–92 and

Figure 1**Three-dimensional structures of 2 S albumins and related proteins**

(a) Wild-type napin Bn1b (PDB entry 1PNB [13]); (b) recombinant pronapin [14]; (c) recombinant RicC3 [15]; (d) nsLTP from maize (PDB entry 1AFH [18]); (e) α -amylase inhibitor from ragi (PDB entry 1BIU [19]); (f) HPS (PDB entry 1HYP [20]).

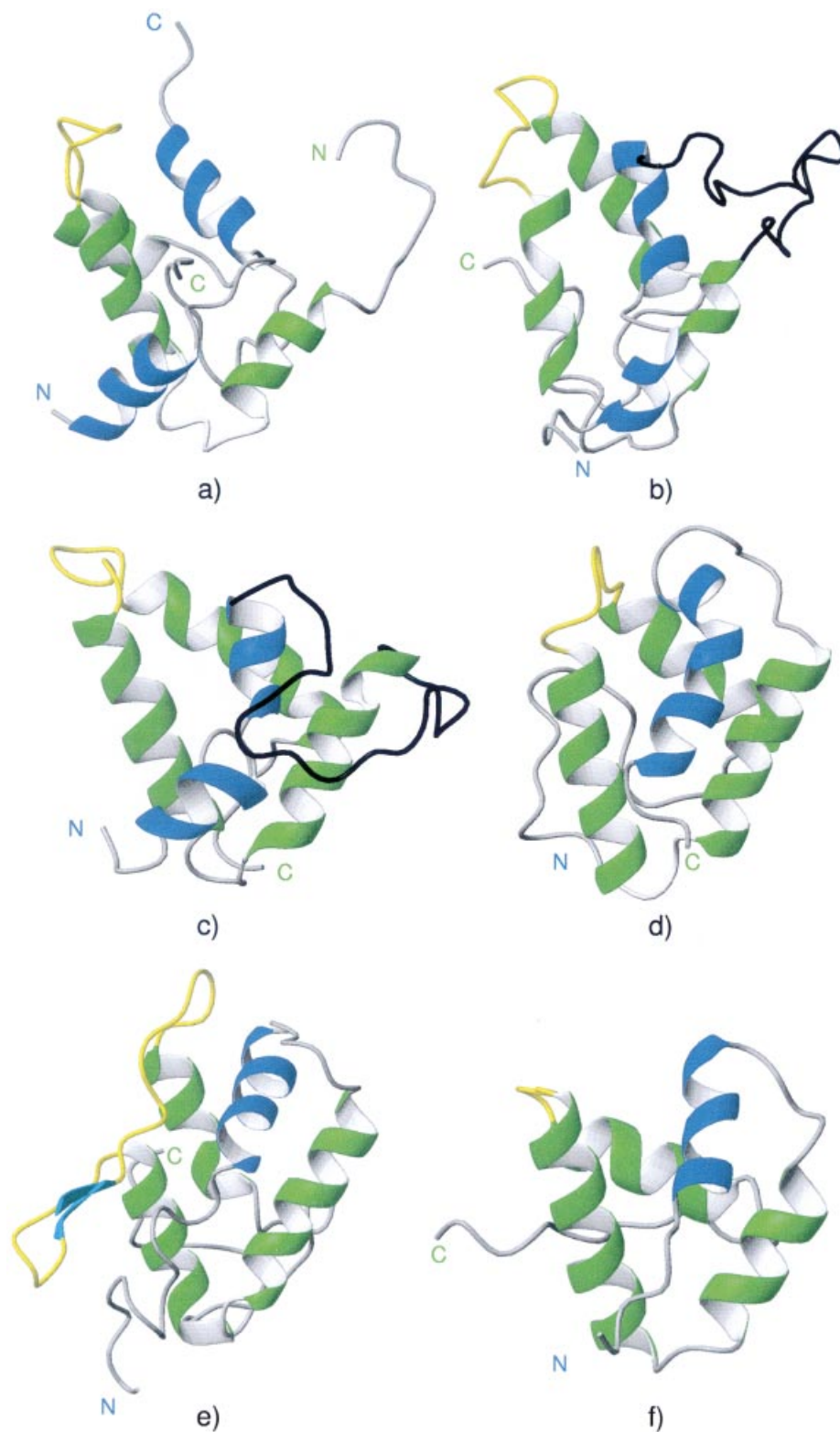
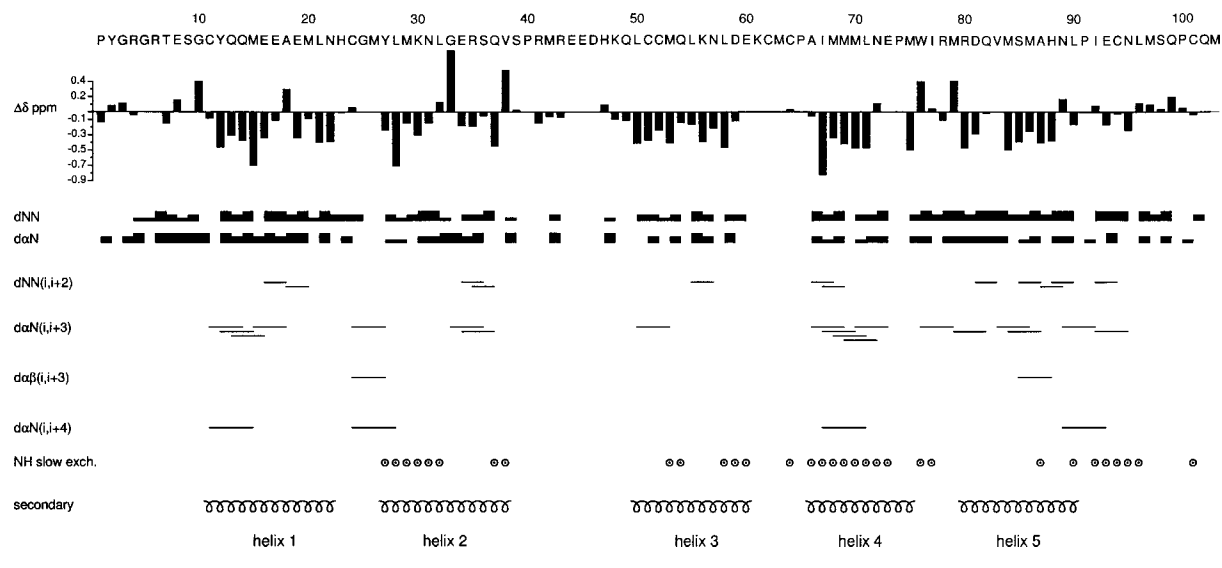


Figure 2

Summary of conformational H α chemical shifts, sequential and short-range NOEs and slow-exchanging backbone amide protons in sunflower SFA-8

A consensus delineation of the five helical regions in SFA-8 is given on the basis of the NMR data.



100–114. The five helices are arranged in a right-handed superhelix, a folding motif first observed in non-specific LTPs (nsLTPs) and detected subsequently in 2 S albumins.

The assignment of the backbone ^1H resonances of SFA-8 is now almost complete ($\sim 90\%$ residues). A large number of NOE cross-correlations have also been assigned, which will form the basis of the three-dimensional structure determination, now in progress. Data currently at hand [H_α conformational chemical shifts, intensity and type of backbone sequential connections, $d_{\alpha\text{N}(i,i+3)}$, short-range NOEs (where $d_{\alpha\text{N}}$ is defined as the NOE connectivity between the C_αH proton and the amide proton resonances), and exchange-protected backbone amide protons] confirm that SFA-8 is an all-helix protein, and allow delineation of its five component helices, which span residues 11–22, 27–38, 50–60, 66–75 and 80–90 (see Figure 2).

Discussion

Figure 1 shows the known structures of 2 S albumins (Figure 1a, napin; Figure 1b, pronapin; Figure 1c, RicC3) and compares them with representative structures of the nsLTP family (Figure 1d; maize [18]), α -amylase inhibitors (Figure 1e; ragi [19]) and HPS (hydrophobic protein from soybean [20]; Figure 1f). Figures 1(a) and 1(b) show 2 S albumins from Brassicaceae. The protein

depicted in Figure 1(a) came from a natural source, and its structure was reported some years ago [13]. Note that the connection between the short chain (in blue) and the large chain (in green) has been cleaved by post-translational events. Figure 1(b) shows the structure of recombinant pronapin, a single polypeptide chain in which the link between chains (in bold) remains intact. Although the structure of this protein is still being refined, an improvement in the definition of its five helical components is already apparent. Also, the spatial arrangement of these five helices is more reliable than that at which we arrived for the native napin. RicC3 (Figure 1c) is the best defined structure we have at present. The positions of helices 1 and 2, as well as the V-shape defined by helices 3 and 4, are very similar to the ones in napin, although the orientation of helix 5 differs markedly. The orientation of helix 5 in pronapin, more towards being orthogonal to helices 3 and 4, is more similar to that in RicC3 than in napin, where helix 5 is ill defined.

The three-dimensional structures of 2 S albumins are very similar indeed to those of the nsLTPs (Figure 1d), α -amylase inhibitors (Figure 1e) and HPS (Figure 1f), which might be expected from the close matching of their corresponding disulphide bond patterns. However, some small differences are apparent. Figure 3 shows an alignment of the sequences of some representative members of the four families (2 S albumin,

nsLTP, α -amylase inhibitor and HPS). The main basis of the alignment is the disulphide bond pattern, complemented by information on the extension and length of the helices. The shaded segments for each family correspond to a consensus of the spanning of the helices in the structures deposited in the Protein Data Bank for each family. As can be seen in Figure 3, the location and extension of the helices approximately coincide in the four families, although the location of the last four helices in the nsLTPs and in HPS is shifted appreciably with regard to those in the 2 S albumins and the α -amylase inhibitors. Of particular note is the shift of helices 3 and 4 in the nsLTPs and of helices 2 and 3 in HPS, which positions the CXC motif in the middle of the latter helix, whereas in the other two families this motif is present at the N-terminus of that helix. There is another important difference between the structure of nsLTPs and those of albumins, α -amylase inhibitors and HSP that is not readily apparent from Figure 1. Helices 3 and 4 in nsLTPs are almost co-planar and helix 5 is much further back than in the other two families, in which the V-shape formed by the homologous helices is more twisted. As a consequence, an internal cavity is present in nsLTPs that is able to host one molecule of a lipid acid, whereas no cavity appears to be present in members of the other families. HPS is more similar to nsLTPs than to 2 S albumins or α -amylase inhibitors.

The molecular architecture described here applies to one of the four structural families into

which more than 40 allergenic proteins with known structure have been classified [21]. The immediate question is: what is the relationship between structure and allergenicity? It is well known that some aspects of protein structure are relevant for allergenicity, such as solubility, compactness, stability, and possibly an ability to interact with lipids in the membrane. These properties are related primarily to transport over mucosal barriers and susceptibility to proteases, so that the proteins can reach their targets intact. In this regard, the structures of the 2 S albumins described here are highly compact, which confers on them a particular stability to thermal denaturation and to digestion by proteolytic enzymes. In fact, the NMR spectra were obtained over a weekend at pH 3.5 and 35 °C, without any sign of denaturation. Also, preliminary thermal differential scanning calorimetry denaturation of a related napin provided values of 100.3 °C and 80 °C for its transition temperature at pH 6 and 3 respectively [22]. This high stability even at low pH values, together with an ability to interact with membranes [23], could underlie the allergenic character of the 2 S albumins, since resistance to digestion and interaction with membranes are key factors for the allergenicity of food components.

From a structural point of view, the major question related to the localization of epitopes able to cross-react with high-affinity antibodies remains to be answered. On epitope mapping of the allergen Sin a 1 from yellow mustard (*Sinapsis alba*) [24], which is sequentially and structurally

Figure 3

Sequence alignment of representative 2 S albumins and members of related protein families

Shown is a sequence alignment of representative members of four related protein families: 2 S albumins (1PNB [13]), nsLTPs (1AFH [17]), α -amylase inhibitors (1BIU [18]) and HSP (1HYP [19]). The spanning of the corresponding helices, shaded in grey, corresponds to a consensus among all members of the families with structures deposited at the Protein Data Bank.

```

1PNB  --QPQKCOREFQQEQH-LRACOQWIRQQLAGSPF
1AFH  ----AISCGOVASA----IAPCISYARG-----
1BIU  --SVGTSCIPGMAIPHNPLDSCRWYVSTRTCGVGP
1HYP  ALITRPSCP-----DLSICLNILGG-----

1PNB  QWGPQQGPSLREQCCNELYQED-----QVCPTLKQAAKSVRVQGG
1AFH  ----QSGSPSAGCCSGVRSLNNAARTADRRACNCLKNAAAGV-----
1BIU  ---RLATDEMKARCCRQLEAIP-----AYCRCEAVRILMDGVVTPS-
1HYP  -----SLGTVDDCCALIGGL-----GDIEAIVCLCIQLRALG-----

1PNB  -----HGPFQSTRIYQIAKNLPNVCNMKQIG-----TCPFIAI---
1AFH  -----SGLNAGNAASIPSKCGVSIPYTISTSTDCSRVN-----
1BIU  GQHEGRLLQDLPGCPPROVQRAFAPKLVTEVECNLATIHGGP---FCLSLLLGAGE
1HYP  -----ILNLNRNLQLIILNSCGRSYPSNA---TCPRT-----

```

related to napins, two immunodominant regions were defined. One is located very close to the 'hypervariable region' of the 2 S albumins, which forms a very flexible loop between helices 3 and 4 (in yellow in Figure 1; see also Figure 3). Recently, a minimal linear IgE-reactive epitope was identified in the major food allergen, Jug r 1, in the 2 S albumin of English walnut [25]. This four-residue epitope, Arg-Gly-Glu-Glu, is also located in the hypervariable region, which reinforces the idea that the major epitope of napins and structurally related proteins may reside in this region. More information is needed, in particular on the structure of the complexes of IgE with 2 S albumins, in order to understand more closely the atomic interactions between antibody and antigen, as well as the role of regions other than the epitope in the correct orientation of the allergen into the interacting region of the Fab fragment of the antibody.

This work has been supported by project PB98-0667 of the DGESIC of the Spanish Ministry of Education and by Industrial Farmacéutica Cantabria S.A. D.P.-U. is supported by a predoctoral grant from the Ministry of Science and Technology. Thanks are given to Mrs C. López for technical assistance. We are grateful to Professor R. Rodríguez, Professor G. Giménez-Gallego and Professor P. Shwery for the generous donation of recombinant pronapin, ¹⁵N-labelled recombinant RicC3 and sunflower SFA-8 respectively.

References

- 1 Youle, R. J. and Huang, A. H. (1978) *Plant Physiol.* **61**, 13–16
- 2 Terras, F. R. G., Torrekens, S., Van Leuven, F., Osborn, R. W., Vanderleyden, J., Cammue, B. P. A. and Broekaert, W. F. (1993) *FEBS Lett.* **316**, 233–240
- 3 Polya, G. M., Chandra, S. and Condron, R. (1993) *Plant Physiol.* **101**, 545–551
- 4 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) *Biotechnology* **7**, 929–932
- 5 Altenbach, S. B., Kuo, C. C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A. and Townsend, J. (1992) *Plant Mol. Biol.* **18**, 235–245
- 6 Bashir, M. E., Hubatsch, I., Leinenbach, H. P., Zeppezauer, M., Panzani, R. C. and Hussein, I. H. (1998) *Int. Arch. Allergy Immunol.* **115**, 73–82
- 7 Varela, J., Navarro Rico, M.-L., Guerrero, A., García, F., Giménez Gallego, G. and Pivel, J. P. (2002) *Methods Find. Exp. Clin. Pharmacol.*, in the press
- 8 Sharief, S. F. and Li, S. S. L. (1982) *J. Biol. Chem.* **257**, 14753–14759
- 9 Menendez-Arias, L., Moneo, I., Dominguez, J. and Rodriguez, R. (1988) *Eur. J. Biochem.* **177**, 159–166
- 10 Monsalve, R. I., Gonzalez de la Peña, M. A., Menendez-Arias, L., Lopez-Otin, C., Villalba, M. and Rodriguez, R. (1993) *Biochem. J.* **293**, 625–632
- 11 Monsalve, R. I., Villalba, M. and Rodríguez, R. (2001) *Internet Symp. Food Allergens* (<http://www.food-allergens.de>) **3**, 57–69
- 12 Pastorello, E. A., Pompei, C., Pravettoni, V., Brenna, O., Farioli, L., Trambaioli, C. and Conti, A. (2001) *Allergy* **5**, 45–47
- 13 Rico, M., Bruix, M., González, C., Monsalve, R. I. and Rodríguez, R. (1996) *Biochemistry* **35**, 15672–15682
- 14 Palomares, O., Monsalve, R. I., Rodríguez, R. and Villalba, M. (2002) *Eur. J. Biochem.* **269**, 2538–2545
- 15 Fernández-Tornero, C., Ramón, A., Navarro, M. L., Varela, J. and Giménez-Gallego, G. (2002) *Biotechniques* **32**, 1238–1242
- 16 Pandya, M. J., Sessions, R. B., Williams, P. B., Dempsey, C. E., Tatham, A. S., Shewry, P. R. and Clarke, A. R. (2000) *Proteins* **38**, 341–349
- 17 Pantoja-Uceda, D., Bruix, M., Varela, J., López-Lucendo, M. I. F., Giménez-Gallego, G., Rico, M. and Santoro, J. (2002) *J. Biomol. NMR* **23**, 331–332
- 18 Gomar, J., Petit, M. C., Sodano, P., Sy, D., Marion, D., Kader, J. C., Vovelle, F. and Ptak, M. (1996) *Protein Sci.* **5**, 565–577
- 19 Gourinath, S., Srinivasan, A. and Singh, T. P. (2000) *Acta Crystallogr.* **55**, 287–293
- 20 Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. and Lehmann, M. S. (1993) *J. Mol. Biol.* **231**, 877–887
- 21 Aalberse, R. C. (2000) *J. Allergy Clin. Immunol.* **106**, 228–238
- 22 Barciszewski, J., Szymanski, M. and Haertle, T. (2000) *J. Protein Chem.* **19**, 249–254
- 23 Oñaderra, M., Monsalve, R. I., Mancheño, J. M., Villalba, M., Martínez del Pozo, A., Gavilanes, J. G. and Rodríguez, R. (1994) *Eur. J. Biochem.* **225**, 609–615
- 24 Menendez-Arias, L., Dominguez, J., Moneo, I. and Rodriguez, R. (1990) *Mol. Immunol.* **27**, 143–150
- 25 Robotham, J. M., Teuber, S. S., Sathe, S. K. and Roux, K. H. (2002) *J. Allergy Clin. Immunol.* **109**, 143–149

Received 19 June 2002