

Allergens of the cupin superfamily

E. N. C. Mills^{*†}, J. Jenkins^{*}, N. Marigheto^{*}, P. S. Belton[†], A. P. Gunning^{*} and V. J. Morris^{*}

^{*}Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, U.K., and [†]Department of Chemical Sciences, University of East Anglia, Norwich Research Park, Colney, Norwich NR4 7TJ, U.K.

Abstract

The cupin family comprises a family of proteins possessing a common β -barrel structure that is thought to have originated in a prokaryotic ancestor. This structural motif is found as a single domain in fungal spherulins, fern sporulins and the germins/oxalate oxidase proteins of plants, while the globular storage proteins of plants, called legumins (11 S) and euvicilins (7 S), are two-domain cupins. The 11 S globulins are hexameric heterologous proteins of $M_r \sim 360\,000$, with each subunit comprising an acidic 30\,000–40\,000- M_r polypeptide that is disulphide-linked to a 20\,000- M_r basic polypeptide. A number of cupins have been identified as major plant food allergens, including the 7 S globulins of soybean (β -conglycinin), peanut (conarachin; Ara h 1), walnut (Jug r 2) and lentil, and the 11 S globulins of peanut (arachin; Ara h 3), soybean (glycinin) and possibly also coconut and walnut. Other members of the cupin superfamily have not been identified as allergens, with the exception of one germin (germination-specific protein) from pepper. Cupins are generally very stable proteins. A summary of our current knowledge of allergenic seed storage globulins will be presented, together with an overview of cupin structure and stability properties, as illustrated by the allergenic soya globulins, glycinin and β -conglycinin.

Introduction

The cupins are a large superfamily of proteins which are thought to have originated by divergent evolution from a common ancestor. They share a common architecture, which has been described as 'double-stranded β -helix' or 'jelly-roll' barrel-like structure. It is from this barrel-like structure that the term 'cupin' (based on the Latin name for a small cask or barrel) was first coined, after

members of this superfamily were found to share the sequence motifs $[G(X)_5HXH(X)_{11}G$ and $G(X)_5P(X)_4H(X)_3N$ (where X is any amino acid residue)] that identified a metal-binding site in many but not all members of this superfamily [1,2]. The basic fold is found in single-domain cupins, such as the fungal spherulins and auxin- and sucrose-binding proteins. In the germins, a diverse family of cupins which have enzymic activity, such as oxalate oxidases and/or superoxide dismutases, the double-stranded β -helix domain is fused to a short all- α -helical domain. The 'bicupins' are thought to have arisen through duplication of this single domain, with the 11 S and 7 S globulin seed storage proteins belonging to this family of cupins.

Seed storage globulins from various legumes and tree nuts make a significant contribution to the human diet, as they are major seed storage proteins, with those from soya being among the most widely consumed. In contrast, globulins from other seeds, such as sunflower, sesame and poppy, are eaten in much smaller amounts. Such patterns of consumption are reflected in the observed patterns of allergy towards cupins. Thus the majority of cupin allergens belong to either the 7 S vicilin-like or the 11 S legumin-like seed storage globulin families, with only a single report of a germin food allergen in the literature [3]. Major food allergens include the 7 S globulins of soya (β -conglycinin [4]), peanut (conarachin; also known as Ara h 1; [5]) and walnut (Jug r 2; [6]) and one of the subunits of the proteolytically processed 7 S globulin of lentil [7]. The 11 S globulins have also been confirmed as allergens in peanut (arachin; Ara h 3; [8]) and soya [9], with preliminary reports of their being allergens in almond [10] and sesame [11]. They have also been implicated as allergens in coconut and walnut [12].

Structure of seed storage globulins

Direct comparison of the protein sequences of 11 S and 7 S globulins shows that they share relatively low similarity, the aligned sequences having identities of around 35–45%. This belies the similarity in their three-dimensional structures, which was first described in detail by

Key words: β -barrel, glycinin, seed storage globulins, soybean, stability.

[†]To whom correspondence should be addressed (e-mail clare.mills@bbsrc.ac.uk).

Lawrence and co-workers [13], who used a structural alignment to identify around 30 residues in the jelly-roll structure that are either conserved or conservatively exchanged across the globulin families. Both the 11 S and 7 S globulins possess two structurally equivalent N-terminal and C-terminal domains, each domain comprising the cupin β -barrel structure, which is illustrated in Figure 1 for the trimeric 11 S globulin precursor from soybean, proglycinin a(1a)B(1b) [14]. While very similar to the three-dimensional structure of the 7 S globulins, a major difference is a significant proportion of disorder (approx. 20%) in the structure of 11 S proglycinin.

Originally identified as salt-soluble proteins by T. B. Osborne in 1924 [15], the globulins occur widely in plant species, having been identified in gymnosperms (conifers) and spermatophytes (ferns) as well as in monocotyledonous and dicotyledonous plant species. The legumin-like 11/12 S storage globulins are multimeric proteins of $M_r \sim 300\,000$ – $450\,000$ that exist as a mixture of trimers and hexamers comprising subunits of $M_r \sim 50\,000$ – $60\,000$ held together by non-covalent interactions. The subunits are the product of a multigene family, of which that of soybean is one

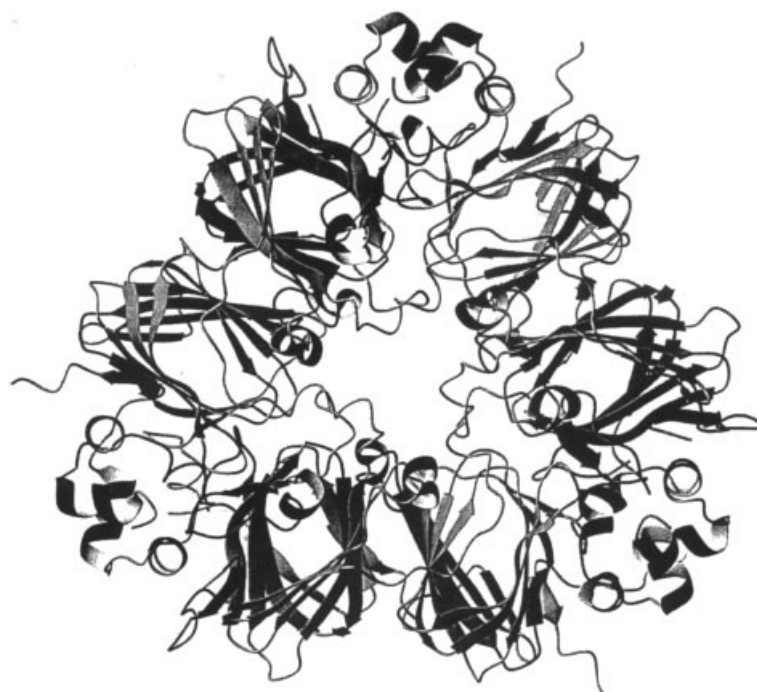
of the best characterized, with around five genes having been identified [16]. They are synthesized as a single polypeptide, which is then cleaved post-translationally to give rise to an acidic ($M_r \sim 30\,000$ – $40\,000$) and a basic ($M_r \sim 20\,000$) polypeptide chain that are linked by a single intermolecular disulphide bond. They are rarely glycosylated, one exception being the major M_r 44 000 acidic subunit of the lupin globulin [17].

The vicilin-like 7/8 S globulins are typically trimeric proteins of $M_r \sim 150\,000$ – $190\,000$, with subunit M_r s in the range $\sim 40\,000$ – $80\,000$, with a typical subunit having an M_r of $\sim 50\,000$. Unlike the 11 S globulins, the 7 S globulins are frequently glycosylated, with one or two N-linked glycosylation sites being located in the C-terminal domain. In the case of the 7 S globulin of peanut, conarachin, this results in a heterogeneous mixture of N-glycans, including $\text{Man}_{5-6}\text{GlcNAc}_2$ and $\text{Man}_{3-4}\text{XylGlcNAc}_2$ [18]. Furthermore, 7 S globulins possessing more than one glycosylation site can exist in several forms, with both singly and doubly glycosylated species having been identified, e.g for the 7 S globulin of green bean, phaseolin [19]. Several 7 S globulins, notably those from pea and lentil, undergo further pro-

Figure 1

Structure of proglycinin, the allergenic 11 S globulin of soybean and a member of the cupin superfamily

The ribbon representation of proglycinin is based on the structure proposed by Adachi et al. [14].



teolytic processing to yield a series of polypeptides of $M_r \sim 12000\text{--}50000$, including the intact protein [20,21]. Despite such processing, these 7 S globulins are still held together by non-covalent forces, retaining the intact $M_r \sim 120000$ trimeric globulin.

The remarkable stability properties of these proteins are described below, using the allergenic soybean globulins as examples, and the role such properties may play in predisposing members of the cupin superfamily to becoming allergens is discussed.

Thermostability and aggregation of globulin allergens

The globulin storage proteins all share a propensity to form large thermally induced aggregates, which form the basis of the widespread utilization of soy protein in foods, as it can form heat-set gel networks [22]. The 11 S globulin of soya, glycinin, will aggregate on heating, forming heat-set gels at concentrations of around 2.5–10 % (w/v), depending on pH and ionic strength. Both 11 S and 7 S globulins, in common with other members of the cupin superfamily, exhibit considerable thermal stability. Thus 7 S globulins have their major thermal transition at around 70–75 °C, while 11 S globulins unfold at temperatures above 94 °C, as determined by differential scanning calorimetry, with the precise values varying between plant species, and with protein concentration and ionic strength. Figure 2 shows the alteration in secondary structure of

10 % (w/v) glycinin at low ionic strength (I 0.08) followed *in situ* during heating, using Fourier-transform IR spectroscopy. As a result of heating, glycinin lost native β -sheet. This is indicated by the decrease in intensity of the amide I band at 1635 cm^{-1} , accompanied by an increase in its helical/random structures, shown by the increase in the ratio of values at 1655 cm^{-1} /1635 cm^{-1} . The formation of intermolecular β -sheet structures is shown by the increase in the amide I band at 1625 cm^{-1} . On cooling there was little change in the intermolecular β -sheet band at 1625 cm^{-1} , while the helical/random content, indicated by the band at 1655 cm^{-1} , decreased, but did not return to that of the unheated material. This indicates that some interactions involved in aggregate formation and unfolding of the glycinin structure were irreversible. However, despite forming a heat-set gel, there was little change in the native-like β -sheet structures, again indicating the structural stability of the cupin fold, although it appears that some local unfolding has to occur to allow the formation of aggregates and hence the heat-set gel network.

Insights are being gained into this process using atomic force microscopy to follow the formation of thermally induced aggregates of β -conglycinin, the 7 S globulin of soya [23]. Aggregates had a fibrous, cylindrical appearance, with a height (diameter) of 8–11 nm, and began to form at temperatures above the main thermal transition temperature for this protein of 75 °C. This coincided with a small change in secondary structure, as indicated by CD spectroscopy, but despite prolonged heating aggregate size did not increase indefinitely, suggesting that certain structural attributes of the β -conglycinin subunits may play an important role in limiting aggregate length. At higher protein concentrations (1 %, w/v), the linear aggregates appeared to form large macro-aggregates, which may be the precursors of protein gel formation (Figure 3).

In addition to disrupting protein structure, heat also causes covalent modifications, of which those involved in glycation and subsequent Maillard rearrangements are among the most important. Such changes may also influence the allergenic activity of globulins. Thus the allergenic activity of the peanut 7 S globulin allergen, Ara h 1, was found to increase following deliberate Maillard modification *in vitro*, as indicated by human IgE binding. These modifications also increased the allergens' thermostability and resistance to digestion, and may underlie the

Figure 2

Fourier-transform IR attenuated total internal reflectance (ATR) spectra of 10 % (w/v) glycinin during *in situ* heating

Glycinin was dissolved in 32.5 mM potassium phosphate buffer, pH 7.6, giving an ionic strength of 0.08 M. The spectra have had buffer background subtracted, and are offset vertically for clarity.

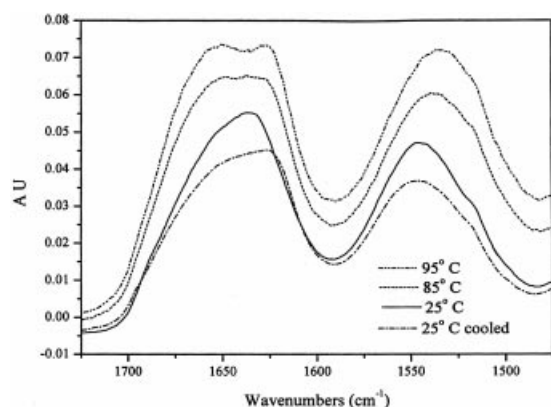
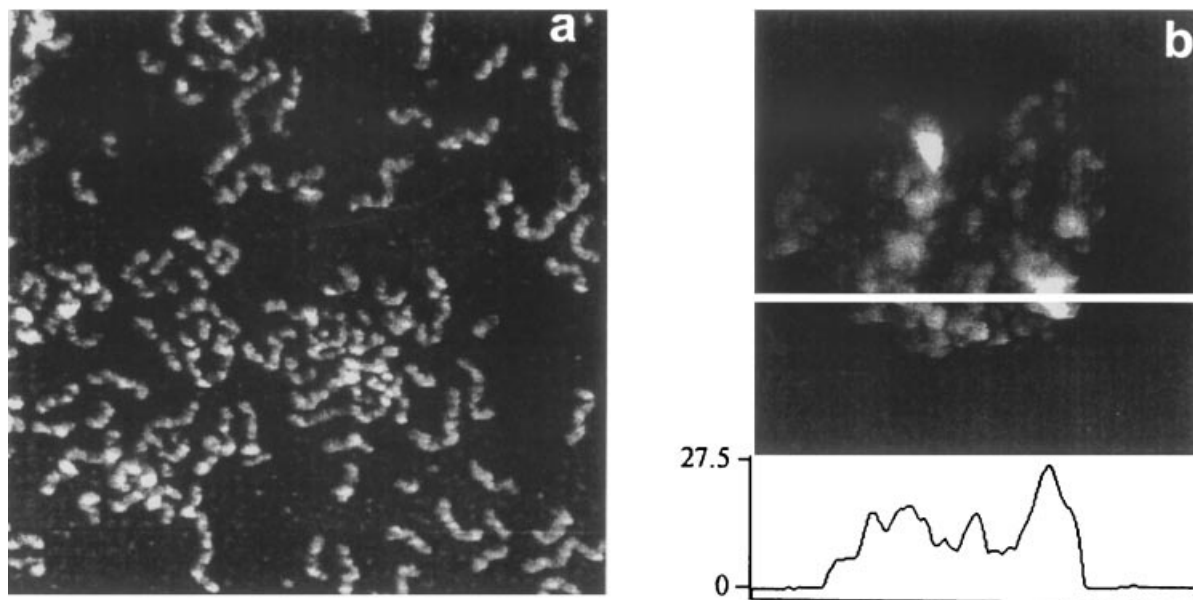


Figure 3

Image analysis of thermally induced β -conglycinin aggregate formation

Data analysis was performed on multiple images obtained for β -conglycinin heated to 100 °C for 10 min at 1/0.2 at (a) 0.4% (w/v) or (b) 1% (w/v). Scan size was 1.2 $\mu\text{m} \times 1.2 \mu\text{m}$. The line profile beneath the image in (b) provides a measure of feature height along the scan line highlighted in the image.



dramatic increase in allergenic activity of peanut extracts from roasted compared with raw nuts [24]. Indeed, it is now apparent that certain methods of cooking peanuts, such as boiling or frying, appear to be more effective at reducing allergenicity, compared with dry roasting [25].

Resistance of globulins to proteolysis

Most of our knowledge regarding the digestibility of storage globulins, and specifically the allergenic globulins from soybean, arises from studies on the utilization of legume protein in relation to animal husbandry. Even when the anti-nutritional effect of lectins and protease inhibitors is taken into account, legume storage globulins are generally less well digested than other sources of dietary protein, such as cow's milk. Thus immunoreactive globulins appear in the ileum 4–6 h after feeding calves on soya flour [26]. The degree of digestion of legume proteins is also dependent on the effect of prior processing (including heat treatment) on the globulins themselves.

Both glycinin and β -conglycinin are susceptible to proteolysis by pepsin [27], although they are partially or fully insoluble between pH 3.5 and 6.5. However, even following proteolysis (with trypsin or chymotrypsin), the 11 S globulin of soya (glycinin) forms stable intermediates of

$M_r \sim 280\,000$, known as glycinin-T or glycinin-C, in which the quaternary structure of the native protein is largely retained. These intermediates result from clipping of the acidic subunits to form fragments of M_r 13 000 and 16 000, with the basic subunits remaining intact [28,29]. Studies using antibodies that recognize proteolysed glycinin indicated that these intermediates may also be formed during digestion *in vivo* in experimental animals [30]. Such stable intermediates may also result from trypsinolysis of 7 S globulins, such as β -conglycinin, with two (M_r 31 550 and 29 500) having been shown to originate from the α/α' subunits and another (M_r 31 500) originating from the β subunit [31].

Cupin structural motifs and plant food protein allergenicity

The double-stranded β -helix that comprises the cupin fold appears to be a remarkably stable structural motif, resisting both thermal denaturation and proteolysis. This stability probably plays an important role in permitting sufficient immunologically active fragments to pass down the gastrointestinal tract, and is responsible, in part at least, for the thermostable nature of the allergenic activity of these proteins. Such proper-

ties, coupled with the abundance of storage globulins in the diet, must contribute to their being able to act as potent allergens. However, while characterized as important allergens in peanut and soybean, the role of cupins in the allergenicity of other nuts and seeds, where they are also abundant components, is less clear. For example, in the Brazil nut, a large proportion of the seed protein is the 11 S globulin excelsin, and yet the major allergen in this plant food species is the 2 S albumin [32]. The poor solubility and stability of these proteins in the dilute salt solutions routinely used to prepare protein fractions for diagnosis may have contributed to them not being identified as allergens in the past. If they are not identified as allergens in the future, this suggests that other factors, hitherto unidentified, may modulate the allergenic properties of cupins in different plant species, either rendering them non-allergenic in foods such as Brazil nut or potentiating their allergenicity in foods such as peanut.

The Institute of Food Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council, U.K. Part of this work was supported by EU FAIR CT-98-4356.

References

- Dunwell, J. M. and Gane, P. J. (1998) *J. Mol. Evol.* **46**, 147–154
- Dunwell, J. M. (1998) *Biotechnol. Genet. Eng. Rev.* **15**, 1–32
- Leitner, A., Jensen-Jarolim, E., Grimm, R., Wuthrich, B., Ebner, H., Scheiner, O., Kraft, D. and Ebner, C. (1998) *Allergy* **53**, 36–41
- Burks, A. W., Brooks, J. R. and Sampson, H. A. (1988) *J. Allergy Clin. Immunol.* **81**, 1135–1142
- Burks, A. W., Williams, L. W., Helm, R. M., Connaughton, C., Cockrell, G. and O'Brien, T. (1991) *J. Allergy Clin. Immunol.* **88**, 172–179
- Teuber, S., Jarvis, K. C., Dandekar, A. M., Peterson, W. R. and Ansari, A. (1999) *J. Allergy Clin. Immunol.* **104**, 1311–1320
- Sanchez-Monge, R., Pascual, C. Y., Diaz-Perales, A., Fernandez-Crespo, J., Martin-Esteban, M. and Salcedo, G. (2000) *J. Allergy Clin. Immunol.* **106**, 955–961
- Burks, A. W., Williams, L. W., Connaughton, C., Cockrell, G., O'Brien, T. J. and Helm, R. M. (1992) *J. Allergy Clin. Immunol.* **90**, 962–969
- Rabjohn, P., Helm, E. M., Stanley, J. S., West, C. M., Sampson, H. A., Burks, A. W. and Bannon, G. A. (1999) *J. Clin. Invest.* **103**, 535–542
- Roux, K. H., Teuber, S. S., Robotham, J. M. and Sathe, S. K. (2001) *J. Agric. Food Chem.* **49**, 2131–2136
- Beyer, K., Bardina, L., Grishina, G. and Sampson, H. A. (2002) *J. Allergy Clin. Immunol.* **109**, S880
- Teuber, S. S. and Petersen, W. R. (1999) *J. Allergy Clin. Immunol.* **103**, 1180–1185
- Lawrence, M. C., Izard, T., Beuchat, M., Blagrove, R. J. and Colman, P. M. (1994) *J. Mol. Biol.* **238**, 748–776
- Adachi, M., Takenaka, Y., Gidamis, A. B., Mikami, B. and Utsumi, S. (2001) *J. Mol. Biol.* **12**, 291–305
- Osborne, T. B. (1924) *The Vegetable Proteins*, 2nd edn, Longmans Green and Co., London
- Nielsen, C. D., Dickinson, T. J., Cho, V. H., Thanh, B. J., Scallion, R. I., Fischer, T. L., Drews, G. N. and Goldberg, R. B. (1989) *Plant Cell* **1**, 313–328
- Duranti, M., Horstmann, C., Gilroy, J. and Croy, R. R. D. (1995) *J. Protein Chem.* **14**, 107–110
- van Ree, R., Cabanes-Macheteau, M., Akkerdaas, J., Milazzo, J. P., Loutelier-Bourhis, C., Rayon, C., Villalba, M., Koppelman, S., Aalberse, R., Rodriguez, R., et al. (2000) *J. Biol. Chem.* **275**, 11451–11458
- Sturm, A., Van Kuik, J. A., Vliegenthart, J. F. and Chrispeels, M. J. (1987) *J. Biol. Chem.* **262**, 13392–13403
- Gatehouse, J. A., Croy, R. R., Morton, H., Tyler, M. and Boulter, D. (1981) *Eur. J. Biochem.* **118**, 627–633
- Casey, R., Domoney, C. and Ellis, N. (1986) *Plant Mol. Cell. Biol.* **3**, 1–95
- Yamauchi, F., Yamagishi, T. and Iwabuchi, S. (1991) *Food Rev. Int.* **7**, 283–322
- Mills, E. N. C., Huang, L., Noel, T. R., Gunning, A. P. and Morris, V. J. (2001) *Biochim. Biophys. Acta* **1547**, 339–350
- Maleki, S. J., Chung, S. Y., Champagne, E. T. and Raufman, J. P. (2000) *J. Allergy Clin. Immunol.* **106**, 763–768
- Beyer, K., Morrow, E., Li, X. M., Bardina, L., Bannon, G. A., Burks, A. W. and Sampson, H. A. (2001) *J. Allergy Clin. Immunol.* **107**, 1077–1081
- Tukur, H. M., Lalles, J. P., Mathis, C., Caugant, I. and Toullec, R. (1993) *Can. J. Anim. Sci.* **73**, 891–905
- Nielsen, S. S., Deshpande, S. S., Hermodson, M. A. and Scott, P. M. (1988) *J. Agric. Food Chem.* **36**, 896–902
- Kamata, Y. and Shibasaki, K. (1978) *Agric. Biol. Chem.* **42**, 2103–2109
- Plumb, G. W., Carr, H. J., Newby, V. K. and Lambert, N. (1989) *Biochim. Biophys. Acta* **999**, 281–288
- Perez, D. M., Mills, E. N. C., Lambert, N., Johnson, I. T. and Morgan, M. R. A. (2000) *J. Sci. Food Agric.* **80**, 513–521
- Kamata, Y., Otsuka, S., Sato, M. and Shubaki, K. (1982) *Agric. Biol. Chem.* **46**, 2829–2834
- Pastorello, E. A., Farioli, L., Pravettoni, V., Ispano, M., Conti, A., Ansaloni, R., Rotondo, F., Incorvaia, C., Bengtsson, A., Rivolta, F. et al. (1998) *J. Allergy Clin. Immunol.* **102**, 1021–1027

Received 19 June 2002