

The Disulphide Mapping, Folding and Characterisation of Recombinant *Ber e 1*, an Allergenic Protein, and SFA8, Two Sulphur-rich 2 S Plant Albumins

Marcos J. C. Alcocer^{1*}, Gareth J. Murtagh¹, Kevin Bailey²
Mireille Dumoulin³, Amparo Sarabia Meseguer¹, Martin J. Parker⁴ and
David B. Archer¹

¹School of Life and
Environmental Sciences
University of Nottingham
University Park, Nottingham
NG7 2RD, UK

²School of Biomedical Sciences
Queens Medical Centre
Nottingham NG7 2UH, UK

³Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge CB2
1EW, UK

⁴Department of Molecular
Biology and Biotechnology.
University of Sheffield, Firth
Court, Western Bank, Sheffield
S10 2TN, UK

We have cloned and expressed genes encoding the allergenic brazil nut 2 S albumin (*Ber e 1*) and the sunflower albumin 8 (SFA8) in the methylotrophic yeast *Pichia pastoris*. We show that both proteins were secreted at high levels and that the purified proteins were properly folded. We also showed that *Ber e 1* is glycosylated during secretion and that the glycan does not interfere with the folding or immunoreactivity. The disulphide map of the *Ber e 1* protein was experimentally established and is in agreement with the conserved disulphide structure of other members of the 2 S albumin family. A model three-dimensional structure of the allergen was generated. During the expression studies and through mutation we have also shown that alteration of the sequences around the Kex2 endoproteolytic processing site in the expressed fusion protein can compromise the secretion by targeting part of the protein for possible degradation. The secreted production of these properly folded sulphur-rich plant albumins presents an opportunity to delineate the attributes that make an allergen and to facilitate the diagnosis and therapy of type I allergy.

© 2002 Elsevier Science Ltd. All rights reserved

*Corresponding author

Keywords: recombinant allergens; plant allergen; brazil nut 2 S; sunflower 2 S; Kex2 processing

Introduction

It has been reported, since the late 1920s that manual workers in the oil milling industry (castor and cotton seed) or in the baking industry ("baker's asthma") became sensitised by inhalation of air-borne particles.^{1–3} During the 1970–1980s the major allergens implicated in those allergies were characterised as belonging to the family of the 2 S albumin/amylase–protease inhibitor proteins (Spies² Youle & Huang⁴ and references therein). Subsequently, the so-called prolamin superfamily was expanded to include other proteins of similar structure, such as lipid transfer proteins (LTP), some prolamins, and indolins, and some members of the superfamily have been implicated as major allergens in a great variety of seeds (wheat,

mustard, sesame, pepper, hazelnut, walnut, peanuts, brazil nut) and fruits (peach, apple, apricot amongst others).^{5–7}

The 2 S albumins are water-soluble proteins present in seeds of a wide range of dicotyledonous and monocotyledonous species.⁸ They are heterodimeric proteins, product of a multigene family comprising subunits of about 30–40 and 60–90 amino acid residues and are synthesised as precursor proteins that are co-translationally transported into the lumen of the endoplasmic reticulum.⁸ After the formation of the four intra-chain disulphide bonds, involving the eight conserved cysteine residues, the folded proteins are transported to the Golgi apparatus, where they are sorted into vesicles for further transport to the vacuole.⁹ The vacuolar targeting signal of brazil nut 2 S albumin in particular (*Ber e 1*) has been mapped and resides at the C-terminal propeptide of four amino acid residues (IAGF).⁹ In plant seeds, during their transport to the vacuole, the

Abbreviations used: SFA8, sunflower albumin 8.

E-mail address of the corresponding author:
marcos.alcocer@nottingham.ac.uk

Ber e 1- 13,582.2 Da

α -mating.....*siaakeegvslekr***▼**EAEAQEECREQMQRQQLSHCRMDMRQQMEESTYQTMPRR**▽**GMEPHMSE
CCEQLEGMDDESCRCEGLRMMMRMMQKEMQPRGRQMRMMRLAENIPSRCNLSPMRCMPGGSIAGFE
*

SFA8- 12,348.0 Da

α -mating.....*siaakeegvslekre***▼**EAPYGRGRTESGCYQMEEAEMLNHCMYLMKNLGRSQQVSPR**▽**MREE
DHKQLCCMQLKNLDEKCMCPAIMMMLNNEPMWIRMRDQVMSMAHNLPICNLMSQPCQMEFPR

Kex-SFA8

α -mating.....*siaakeegvslekr***▼**APYGRGRTESGCYQMEEAEMLNHCMYLMKNLGRSQQVSPR**▽**MREEDHKQ
LCCMQLKNLDEKCMCPAIMMMLNNEPMWIRMRDQVMSMAHNLPICNLMSQPCQMEFPR

Figure 1. Expressed sequences. Expressed 2S albumin sequences showing: mature expressed protein (capital letters), *S. cerevisiae* α -mating factor peptide (italic lower case), Kex2 processing site (bold italic lower case), non-secreted peptide (italic lower case underlined), main (▼) and secondary (▽) proteolytic processing sites, experimentally determined N-terminal sequence (underlined) and amino acids not present on the mature plant native protein (italic capital). The serine residue underlined by * was mutated to methionine in the construct S- > M.

2 S pre-proteins are processed¹⁰ to two subunits, in the case of *Ber e 1* to units of 3628 Da and 8500 Da, with loss of short linker and flanking sequences in at least three stepwise cleavages from the precursor protein.^{11–13} Variant types of 2 S albumins, such as the sunflower 8 2 S albumin (SFA8) described here, are not cleaved *in planta* into large and small subunits but remain as a single monomeric subunit.^{14,15} Although some functions have been assigned to some 2 S albumins, their specific biological roles are mostly unknown.⁸

The allergenicity of *Ber e 1* is well-established.¹⁶ Historically, *Ber e 1* was the first food allergen transferred by transgenic techniques from one plant to another, triggering strong international debate.¹⁷ In contrast, the allergenicity of SFA8 is disputed. While one group has described allergic patients that specifically recognised 2 S protein from the sunflower seed,¹⁸ the large consumption of this seed in countries like Spain, Germany and Greece, and the low incidence of reported specific cases of allergic individuals, suggest otherwise.

In type I allergy there appears to be a breakdown of tolerance to innocuous proteins that is manifested by an immune response that causes a direct hypersensitivity reaction. The cross-linking of mast cell-bound immunoglobulin E (IgE) antibodies by allergens represents the signal for the release of inflammatory mediators. This cascade of events relies fundamentally on antibody recognition of the allergen and the specific “help” of T lymphocytes that are stimulated by linear fragments of the antigen.¹⁹ Although several linear IgE binding epitopes obtained by overlapping peptides are known, including some from 2 S albumins,^{20,21} antibodies generally recognise conformational epitopes brought together by folding of the peptide chain.²² These complementary functions, conformation specificity of the allergen–antibody and the recognition of the linear peptides by T cells, have been successfully exploited in the manipu-

lation of allergens in the specific immunotherapy treatment (SIT) of patients suffering from type I allergy.²³

In order to be able to manipulate and study the three-dimensional structure of *Ber e 1* and SFA8 as representatives of the 2 S family of allergens, and maintain the conserved pattern of cysteine found in the native seed protein, we directed the expression of both 2 S albumins as fusion products to the secretion pathway of the methylotrophic yeast *Pichia pastoris*. Here we show that both proteins were secreted at high levels and that both proteins were properly folded. The post-translational processing and disulphide map of the brazil nut protein were established and a model three-dimensional structure of the allergen was generated. Through mutation we have also shown that alteration of the sequence around the Kex2 endoproteolytic processing site encoded by the expression construct can compromise the secretion by targeting part of the protein to possible destruction.

Results and Discussion

Design and cloning

Aiming at directing the product proteins to the secretion pathway of *P. pastoris*, the mature sequences of the 2 S albumin genes were fused to the *Saccharomyces cerevisiae* α -mating sequence during the construction of the expression cassettes. As shown in Figure 1, the bases encoding the amino acids EAEA after the processing site (KR) were maintained to avoid inaccuracy in the final separation of the fusion α -mating peptide:albumin by the Kex2 peptidase. The shuttle pPIC9 vector containing the albumin sequences was transformed into *P. pastoris* and selected for histidine auxotrophic mutants. In pPIC9 the α -mating gene

is under the control of the AOX1 promoter, providing induction of expression by the presence of methanol in the media.

All clones produced recombinant 2S proteins and higher than average protein-producer clones (4–8% of all transformants) were identified by dot-blot screening. In order to avoid complications during protein purification and to facilitate future labelling experiments, only minimal medium was used throughout expression studies. Under these conditions, the clones best expressing *Ber e 1* and SFA8 yielded around 200 mg and 80 mg of the specific albumin per litre respectively, after 72 hours of induction at 30 °C.

Protein concentration and purification

After the cell separation and concentration of proteins, albumins from the supernatant were purified by a single passage through a heparin-Sepharose column and size fractionation as described below. The heparin column was unusually specific for the purification of the 2S albumins. Under the conditions described, it was possible to concentrate and purify both 2S albumins described here and also a variety of other 2S proteins derived from plant seeds or expressed in *Pichia* (unpublished results).

Antibody recognition

Both recombinant proteins were specifically recognised by rabbit polyclonal antibodies raised against the native seed protein as shown in Figure 2. Moreover, both recombinant proteins showed the same pattern of cross-reactivity of the native seed protein in an ELISA format when sera from nut-allergic subjects were used (unpublished results).

Kex2 processing

Kex2 and related proteases are the main enzymes responsible for the processing of secreted proteins in eukaryotic systems. They are located within the secretory system and cleave following

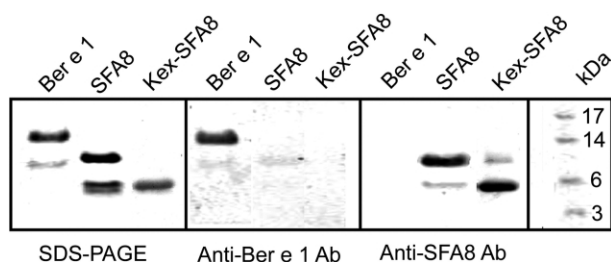


Figure 2. SDS-PAGE-Western blot. Recombinant *Ber e 1*, recombinant SFA8 and the product of Kex-SFA8-construct protein expression, in which the Kex2 site was mutated, separated by SDS-PAGE and developed using specific rabbit polyclonal antibody anti-native *Ber e 1* or anti-native SFA8 protein.

exposed dibasic amino acid residues (e.g. Lys-Arg) during the transit of proteins to the cell exterior. As shown in Figure 2, both *Ber e 1* and SFA8 proteins were produced as main products with a nick in approximately 20% of the proteins at a secondary processing site downstream of the principal protease processing site. The amount of nicked product was variable amongst different batches of purified proteins. The N-terminal amino acid sequence results (Figure 1) showed that the engineered processing site (KR) was mainly recognised and cleaved in *P. pastoris* and that the presence downstream of charged residues PRR and SPR in *Ber e 1* and SFA8, respectively, acted as secondary processing sites. The same pattern of cleavage was observed in early stages of secretion (< five hours) when proteins were synthesised in the presence of [³⁵S]methionine (unpublished results), suggesting that processing occurred intracellularly, possibly at later stages of the secretion, and is not due to the degradation of the proteins by secreted proteases in the medium.

An attempt to improve the trimming of the SFA8 protein by shortening the N-terminal amino acid sequence added to SFA8 by three residues (Kex-SFA8, Figure 1) unexpectedly resulted in the extensive deletion of the N-terminal subunit sequence (Figure 2).

The sequence around the dibasic amino acid Kex2 cleavage site is known to influence the extent and fidelity of cleavage in other fungi, including yeasts.^{24–27} Interestingly though, only trace amount of the intact SFA8 protein and none of the fused product alpha mating type :: SFA8 N-terminal sequence were detected in the extract when the anti-SFA8 polyclonal antibody was used in a Western format. Although the specificity of the polyclonal antibody for the SFA8 N-terminal sequence might explain this result, the absence of any band in SDS-PAGE suggests that the N-terminal alpha-mating peptide :: SFA-8 fusion is degraded intracellularly and not secreted. Whilst unusual, similar results were observed during the processing of the prepro- α -factor in *S. cerevisiae*, which the signal and glycosylated prepro sequence were not removed into the ER and possibly internalised after the processing in the secretory vesicles.²⁸ Our studies with the plant albumins confirm the importance of context for efficient cleavage of the Kex2 processing site.

Glycosylation

Possibly due to their compact structure and disulphide cross-linking, the 2S albumin proteins have shown a characteristic resilience to ionisation in mass spectra (MS).

When the recombinant SFA8 protein was analysed by electro-spray (Q-ToF), there was no evidence of attached sugars as the main peak possessed exactly the predicted mass value (Figure 3) when deconvoluted to singly charged ions. Longer (>72 hours) expression times, however,

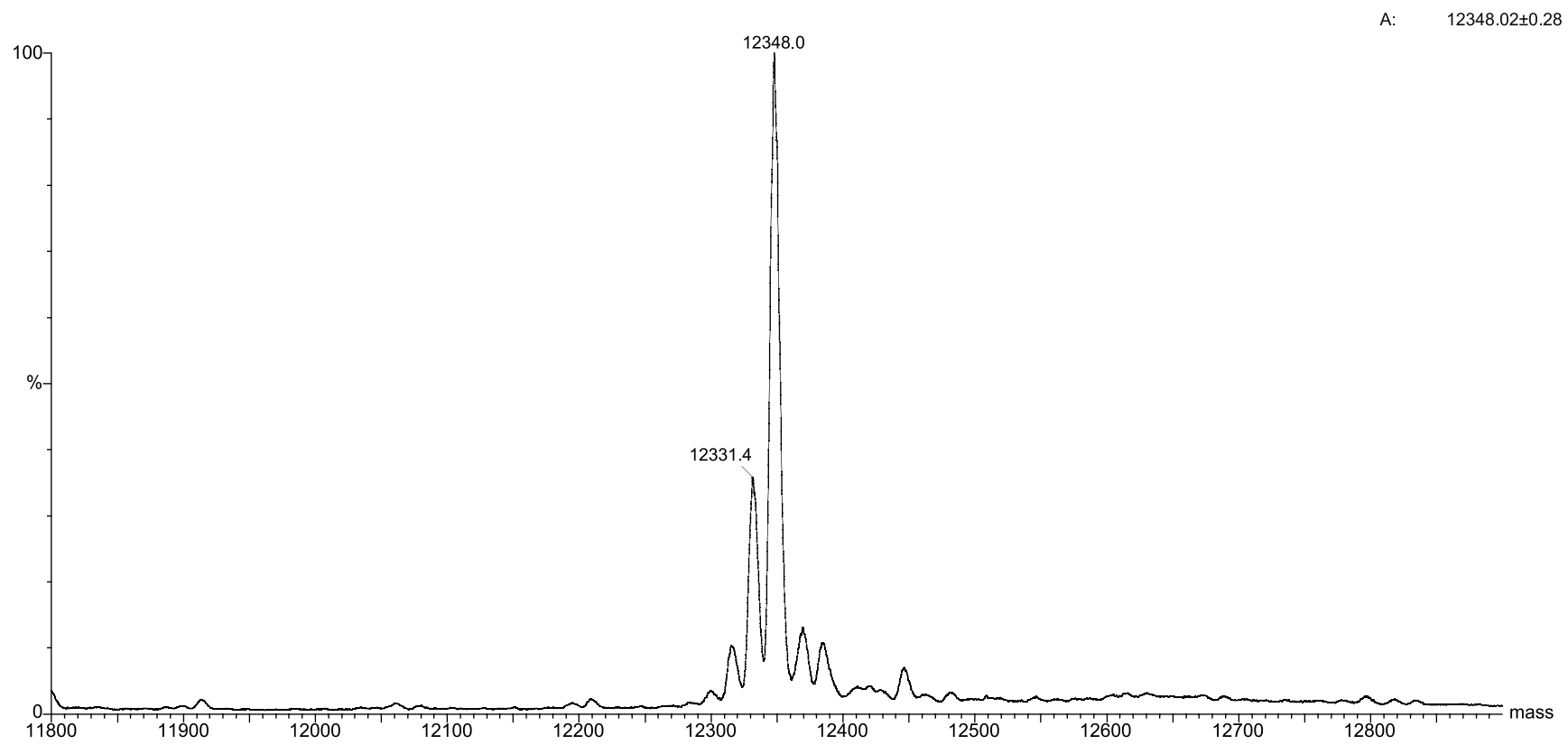


Figure 3. Q-TOF mass spectra of sunflower 2 S albumin 8 (SFA8).

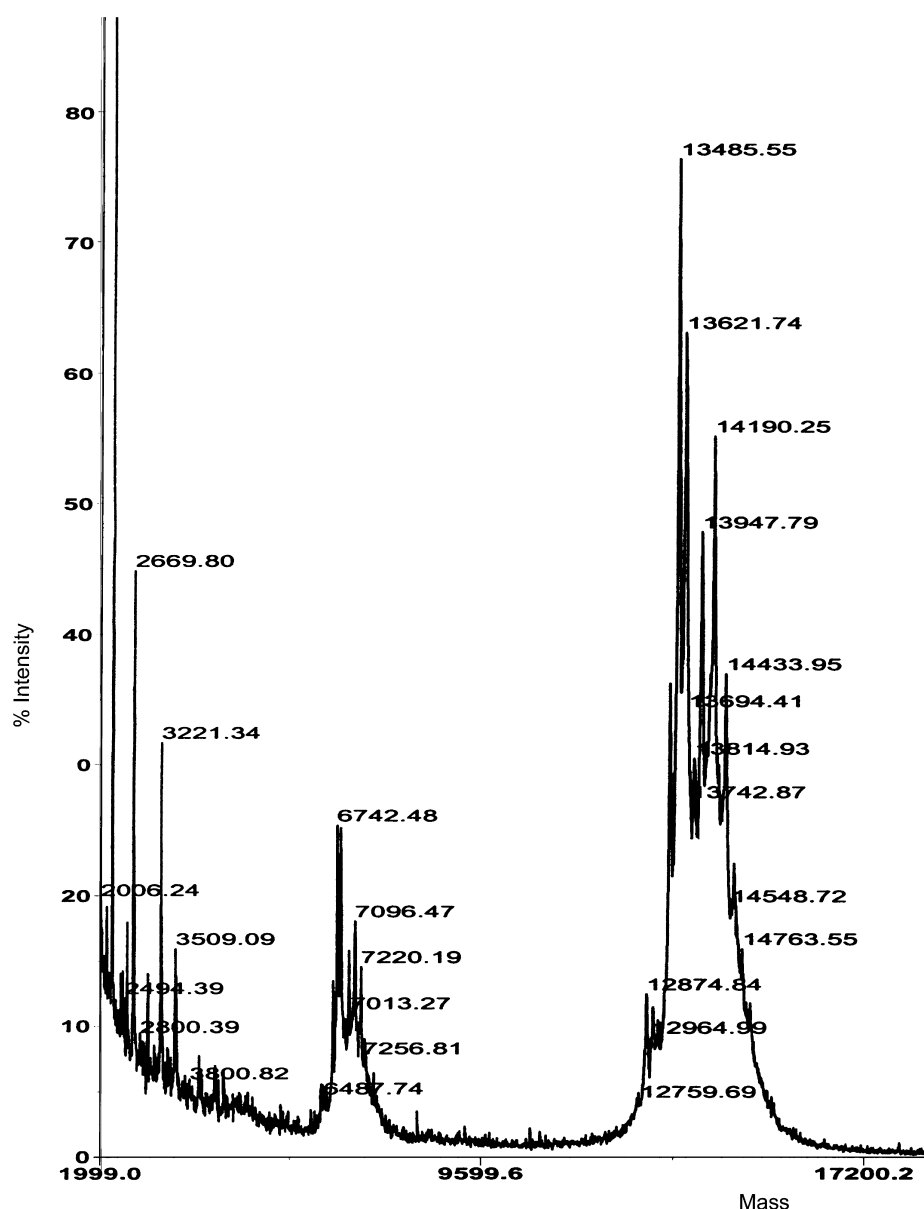


Figure 4. MALDI-TOF mass spectra of the recombinant brazil nut 2S albumin (*Ber e 1*).

resulted in extra peaks with added mass of 16 Da (not shown), suggesting that oxidation of methionine residues was occurring during the high aeration conditions employed during expression. Due to the high methionine content of some 2S albumins, oxidation of methionine to methionine sulphoxide was expected and also reported in early studies as artefacts during extraction from plant seeds.¹¹

After several unsuccessful attempts with the electro-spray, the mass of the *Ber e 1* protein was finally ascertained using a MALDI-TOF system. Instead of the sharp peaks observed for SFA8, *Ber e 1* showed a pattern typical of a glycosylated protein with a much broader mass distribution (Figure 4). The presence of attached sugars was also established by 1D NMR (M. Williamson, personal communication) and further confirmed

by the binding of *Galanthus nivalis* agglutinin to terminal mannose (data not shown). No change in mobility of the recombinant *Ber e 1* protein was observed in SDS-PAGE following treatment with either *N*-glycosidases H or F, or *O*-glycosidase.

The mass results and the undetectable gel-shifts following treatment of the recombinant *Ber e 1* with *N* and *O*-glycanases indicate that the saccharides attached to the protein are small. Mutation of the serine to methionine (S → M, Figure 1) at the potential *N*-glycosylation site (N-X-S/T, where X represents any amino acid residue) in the *Ber e 1* did not abolish binding of the *G. nivalis* agglutinin or alter the overall MALDI-TOF spectrum, suggesting that *O*-type glycans might be involved. The absence of *N*-glycosylation in the recombinant protein is consistent with the lack of evidence for

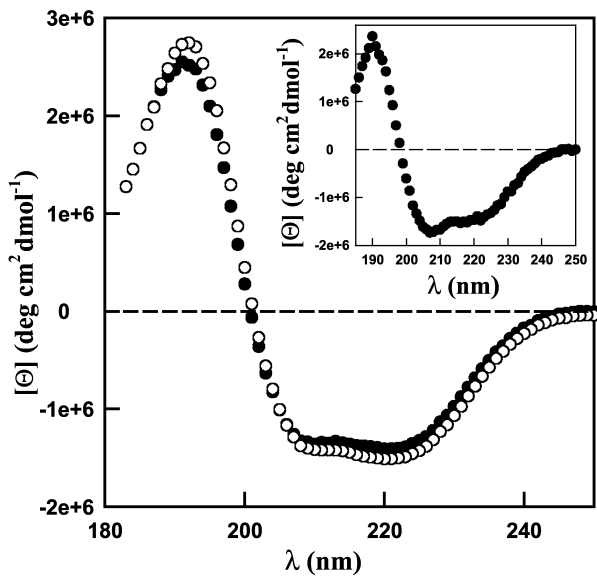


Figure 5. CD spectra in the far UV region of *Ber e 1* (○), recombinant *Ber e 1* (●) and recombinant SFA-8 (inset) at pH 6.8 and 25 °C.

the attachment of N-linked glycans to the native protein derived from plant seed.¹¹

Different degrees of glycosylation of allergenic proteins produced from *P. pastoris* have been reported^{29,30} and reviewed elsewhere.^{31,32} In particular, short O-linked saccharides of mannose containing alpha 1,2 glycosidic or phosphorylated mannose linkages have been characterised.^{31,33} The published results indicate that the short glycans added by *P. pastoris* did not alter the 3D structures³¹ or the immune reactivity^{34,35} of expressed proteins but did reduce the authenticity of product.³⁶

Circular dichroism

In contrast to the recombinant *Ber e 1*, the recombinant SFA8 protein had a molecular mass identical with the expected value, indicating that glycosylation was not a problem. Furthermore, the CD spectra (Figure 5 inset) showed virtually identical traces to the native protein isolated from seed and published elsewhere.³⁷ Whilst the comparison of recombinant SFA8 and the published native seed protein was straightforward, the presence of short glycans attached to the recombinant *Ber e 1* and their influence on protein folding needed to be addressed. The CD spectra of the native seed and the recombinant *Ber e 1* protein were compared (Figure 5). Despite processing of the two subunits and loss of the short linker and flanking sequences in at least three stepwise cleavages in the wild-type protein, the spectra of the native and recombinant *Ber e 1* proteins recorded at 25 °C and pH 6.8, are very similar to each other (Figure 5), suggesting that the recombinant and native proteins have similar secondary structure.

In order to compare the physical characteristics of the recombinant and native proteins, the stability of *Ber e 1* and the recombinant *Ber e 1* were further tested against pH, temperature and chemical denaturants. The circular dichroism (CD) spectra of the native and recombinant proteins were very similar to each other at different pHs. For the two proteins, spectra obtained at pH 2.2 and 25 °C (Figure 6) were superimposable to those recorded at pH 6.8 (Figure 5). Temperatures over 62 °C were necessary (Figure 6 inset) to start unfolding both *Ber e 1* proteins at pH 2.2, further demonstrating the high resistance of these proteins to acidic pH. At 95 °C, both proteins were significantly destabilised although the transition was not complete. Upon cooling, the denaturation was

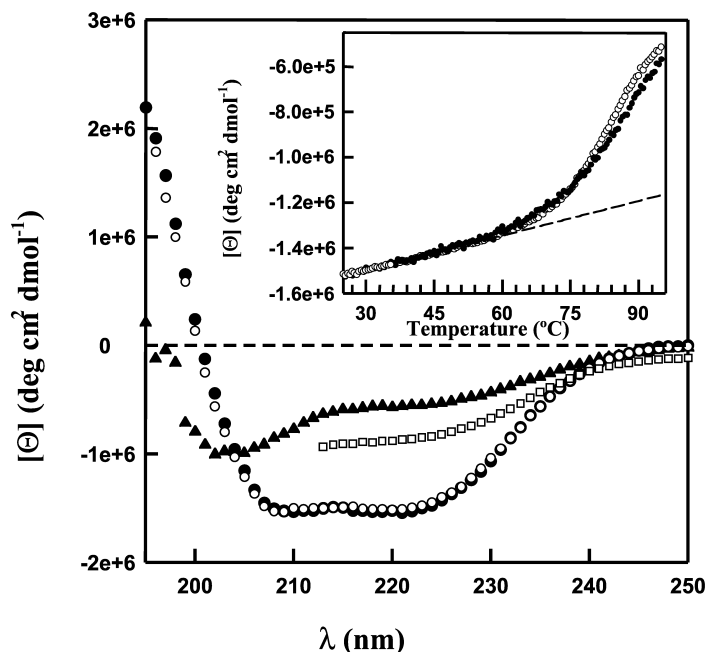


Figure 6. CD spectra of recombinant *Ber e 1* under various conditions in 10 mM glycine buffer pH 2.2. (●) 25 °C upon heating; (▲) 95 °C; (○) 25 °C upon cooling; and (□) 6 M guanidinium chloride. The CD spectra of the native *Ber e 1* are similar to the recombinant protein under the same conditions and are not shown for clarity. Insets: heat induced unfolding transition of *Ber e 1* (○) and *rBer e 1* (▲) followed by far UV CD measurements at 222 nm. The dotted lines represent the slope of the pre-transition.

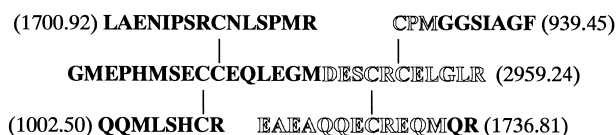


Figure 7. Disulphide mapping of the recombinant *Ber e 1* protein. Derived monoisotopic mass ($M +$) is shown between brackets. In bold and outlined are the products determined by cyanogen bromide cleavage.

largely reversible. The two proteins were also very resistant against chemical-induced unfolding. No changes in the secondary structure were observed up to a concentration of at least 4.5 M guanidinium chloride and the protein exhibited a significant amount of secondary structure in 6 M guanidinium chloride (Figure 6). A comprehensive comparison of the proteolytic, thermal, acid and chemical stability of the recombinant and native *Ber e 1* protein will be presented elsewhere. These results demonstrate that the recombinant *Ber e 1*, albeit being glycosylated, has the same properties as the native seed protein.

Disulphide mapping

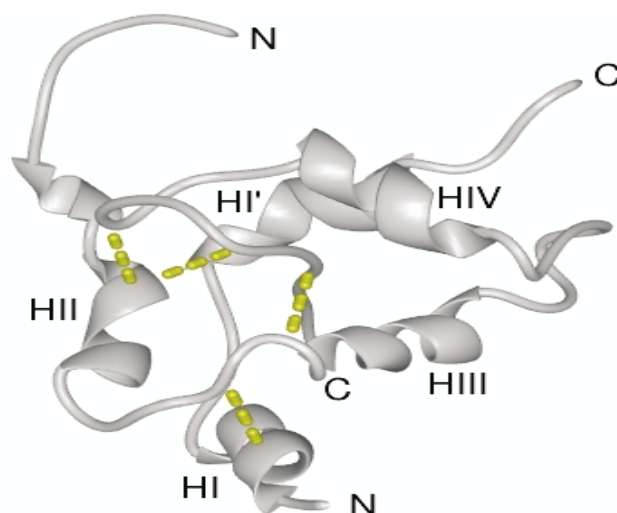
One of the main features of the 2S family of plant proteins is the conserved structure of the disulphide bonds.³⁸ Therefore, determination of the disulphide connectivities is an essential experi-

mental first step towards modelling and elucidating the full protein structure. In order to confirm the CD results and ascertain whether the *Ber e 1*, despite being glycosylated, is properly folded, disulphide mapping was undertaken. A number of peaks were resolved from the reverse phase HPLC analysis of *Ber e 1* tryptic digests. A relatively abundant peak eluting after 30.81 minutes was collected for further analysis. Mass spectrometry revealed the presence of an apparently abundant ion of 8330.13 mass units when deconvoluted to a singly charged species. N-terminal sequencing suggested that the fragment was composed of a number of peptides, linked by disulphide bridges. Following reduction with DTT, several new ion species were detected. MS-MS fragmentation results (Figure 7) indicated that the amino acid sequences present in the complex corresponded to the predicted *Ber e 1* disulphide core and are in agreement with the N-terminal sequencing of the large complex prior to reduction. In order to precisely map the position of the disulphide bridges, the 8330.13 mass unit peptide fragment was further digested with cyanogen bromide. Mass spectrometry revealed the presence of an apparently abundant series of ions of 2848.64, 2866.65, 2876.62 mass units when deconvoluted to singly charged species. The N-terminal sequencing of this peak (Figure 7) agreed with the predicted mass of the cyanogen bromide fragment. These results are in agreement with the disulphide map determined for other members of the 2S albumin family³⁸ and

(a) Sequence alignment.

```

1PNB  QPQKQREFOQEHLRACQOWIRQQLAGSPFQSGPQOG-PWLREQCCNELYQEDQVVCVPTLKQAAKSVRVQGGHGFQSTRIYQ-IAKNLPNVCNMKQIGTCFFIAI
      + C + Q + Q L C +RQQ+ S P G +CC +L D+ C C L+ + + R +A N+P CN+ + CP
Ber e 1 QEECREQMQRQQMLSHCRMDMRQQMEES----MPPRGMEPHMSECCQELEGMDSECRCEGLRMMRMMQKEMQPRGEQMRMRMLAENIPSRCNLSPM-RCPMGGS
  
```



(b) Model structure.

Figure 8. Molecular modelling of *Ber e 1* protein. (a) Sequence alignment between 1PNB and *Ber e 1* demonstrating the absolute and conservative (+) amino acid identities. (b) Ribbon representation of the backbone of the model *Ber e 1* structure. The disulphide bonds are represented as broken lines. Picture produced using MidasPlus (UCSF Computer Graphics Laboratory).

together with the CD results strongly suggest that the recombinant *Ber e 1* is properly folded.

Modelling

On the basis of the conserved 2S albumin disulphide connectivity determined for the *Ber e 1* protein, together with both sequence and structural alignments with other 2S albumin proteins, it was possible to construct a model for this protein using the structure of the napin seed storage protein³⁹ (1PNB) as a template. The experimentally determined structure of 1PNB is that of the mature, proteolytically cleaved protein, hence the model generated here for *Ber e 1* corresponds to the proteolytically cleaved form. The sequence alignment used for building the *Ber e 1* model is shown in Figure 8(a). This alignment exhibits ~26% absolute and ~40% conservative amino acid sequence identity. The final, energy-minimised model for *Ber e 1* is shown in Figure 8(b), and exhibits acceptable geometrical parameters, assessed using the program Procheck.⁴⁰ The quality of the alignment and structural model allow us to classify *Ber e 1* as a member of the “four helices; folded leaf; right-handed superhelix; disulphide-rich” fold family with reasonable confidence.

In conclusion we have shown that the allergenic *Ber e 1* and SFA8 2S albumins were secreted at high levels from *P. pastoris*. Through mutation, we also showed that altered sequence around the Kex2 processing site can compromise secretion, probably by targeting the fusion protein to destruction. We also showed that both proteins were properly folded and that the recombinant *Ber e 1* is glycosylated, probably by O-linked short saccharides. The disulphide map of the recombinant *Ber e 1* brazil nut protein was established and is in agreement with the map determined for other members of the 2S albumin family. Using those data a model structure of the allergen was generated and is consistent with the CD spectrum results. The secreted production of these properly folded sulphur-rich plant allergenic albumins will facilitate the determination of the 3D structure of this important family of allergenic proteins and presents an opportunity to delineate the attributes that make an allergen.

Material and Methods

Media composition for *P. pastoris*

P. pastoris GS115 (Invitrogen) was maintained in YPD broth (10.0 g yeast extract (Difco), 20.0 g peptone, 20.0 g dextrose per litre). MD plates (13.4 g yeast nitrogen base, 10.0 g glucose, 0.4 mg biotin, 100 ml 1 M K₂HPO₄/KH₂PO₄ (pH 6.0), 20 g agar per litre) were used for plasmid selection. BMG broth (13.4 g yeast nitrogen base, 0.4 mg biotin, 10 ml glycerol, 100 ml 1 M K₂HPO₄/KH₂PO₄ (pH 6.0) per litre) was used for enrichment. BMM broth (13.4 g yeast nitrogen base, 0.4 mg biotin, 5 ml methanol, 100 ml 1 M K₂HPO₄/KH₂PO₄ (pH 6.0)

per litre) was used for induction. Unless otherwise stated, all the chemicals were supplied by Sigma.

Construction of expression plasmids

A plasmid containing the cDNA sequence of sunflower albumin 8 (Genbank X56686) was kindly donated by T. J. Higgins (CSIRO Plant Industry, Australia). The brazil nut genomic DNA was isolated from seed using a plant DNA extraction kit (Nucleon Phytopure, Amersham) according to the manufacturer's instructions. The 2S albumin genes were amplified by PCR using the primers: 5'-TCTCTCGAGAAAAGAGAGGCTGAAGCTCCCTATGGCCGAGGTAGAACTGA and 5'-CTGAATTCTCATTACATTGGC ATGGTTGGGACAT; 5'-GAAGGGGTATCTCTCGAGAAAAGAG AGGCTGAAGCTATGGGTGT CTACACATTTGAAAACGA and 5'-ATTGAATTCA GAACCCGGCAATGGAGCCAC; for SFA8, and *Ber e 1*, respectively, at 35 cycles of 94 °C 30 seconds, 54 °C 30 seconds, 72 °C 30 seconds, using Amplitaq (two units/100 µl; Applied Biosystems) according to the manufacturer's instructions. The PCR products (ca 400 bp) were gel-purified, digested with *Eco*RI/*Xho*I enzymes for four hours at 37 °C and ligated into pPIC9 (Invitrogen) previously digested with the same enzymes. The resulting plasmids were transformed into *E. coli* XL1Blue (Stratagene) and selected under ampicillin (100 µg/ml) following standard protocols.⁴¹ Plasmids containing the inserted gene were purified (midi kit, Qiagen) and sequenced using the Sanger method with dye terminators (Applied Biosystems). The small intron (70 bp) from genomic *Ber e 1* (Genbank X54491) was removed by Splice over extension PCR, cloned as described above and re-sequenced.

Transformation and expression in *P. pastoris*

pPIC9-derived plasmids containing the 2S protein-encoding sequences were linearised at the *Sac*I site for four hours at 37 °C. The linearised plasmids were then transformed into *P. pastoris* GS115 (5 µg/transformation) and the transformed strains were selected on MD plates following electroporation according to the manufacturer's instructions (Invitrogen). Clones able to grow in the absence of histidine were inoculated into a sterile 96-well plate format containing BMG and incubated overnight at 30 °C. The 96 clones were then expanded to 4 × 24-well plates containing BMG (1.5 ml/well) and incubated overnight at 30 °C. After centrifugation (750g for 15 minutes), the 24-well plates were emptied by suction and the BMM broth (containing methanol at 0.5% v/v) added. Expression was allowed to take place in an orbital shaker (200 rpm) for 48 hours with addition of methanol (0.5% v/v) every 12 hours.

Dot-blot screening

After induction with methanol, the 4 × 24-well plates were centrifuged (750g for 15 minutes) and the supernatant (500 µl) transferred under vacuum to a wetted PVDF (Millipore) membrane in a dot-blot format system. The membrane was then blocked by incubation with 10 ml of BSA (bovine serum albumin, 5% (w/v)) in TBS (Tris buffered saline, 20 mM Tris (pH 7.5), 0.9% (w/v) sodium chloride) for one hour. All incubations were performed at 37 °C in a standard hybridisation oven. After three washes (five minutes each) with TBST (TBS plus Tween 20, 0.1% w/v), the membranes were incubated

for two hours with the specific rabbit polyclonal antibody anti-native *Ber e 1* or anti-native SFA8 diluted (1:1000) in TBST plus 20% (v/v) of *P. pastoris* supernatant extract that was previously prepared and did not contain the target protein. The membrane was then washed three times with TBST and incubated for a further one hour with the secondary alkaline phosphatase-labelled anti-rabbit antibody, diluted 1:20,000 in TBST. After the three washes with TBST the membrane was incubated with Nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate toluidine substrate. Once the reaction was completed, the blot was washed with distilled water and dried.

Large-scale production

Selected clones were grown at high density ($A_{600\text{ nm}}$ 3–6) in five \times 4 l-flasks containing BMG (2 l) at 30 °C. The culture was then centrifuged (6000g for 15 minutes), and the pellet re-suspended in BMM (10 l) and incubated with agitation (200 rpm) at 30 °C for 48–72 hours with addition of methanol (0.5% v/v) every 12 hours. After induction, the supernatant was separated from cells by centrifugation (6000g for 15 minutes) and filtered through a 0.45 μm membrane. The supernatant was then concentrated (50 \times) and the salt eliminated by buffer exchanged (citrate buffer, 10 mM, pH 4.3) using a tangential flow system (Viva Science, cut-off 5000 M_r). The concentrated sample was then loaded on to a heparin-Sepharose column (3 \times 5 ml) fitted to a FPLC (Pharmacia) system and the 2 S proteins eluted at 400–600 mM sodium chloride. The eluted peak was then loaded on to a Sephadex 75 column (Pharmacia, 2.5 cm \times 90 cm) run (2 ml/minute) with phosphate buffer saline (100 mM sodium phosphate (pH 7.5), 1 M NaCl) and the purified 2 S protein eluted after two hours. The eluted peak was dialysed overnight against water (4 °C) and freeze-dried.

SDS-PAGE/Western blot

SDS-PAGE and Western analysis were performed following the manufacturer's instructions (NOVEX—Invitrogen), using Immobilon PVDF membrane (Millipore) and antibody dilutions and incubations as described above for the dot-blot screening.

Glycan analysis

The glycosylation of the proteins was confirmed by the binding of *G. nivalis* agglutinin to terminal mannose in a Western blot format following the instructions from the glycan differentiation kit (Boehringer Mannheim Biochemica). De-glycosylation was carried out following the manufacturer's instructions using peptide-*N*-glycosidase F or endoglycosidase-H, after heat-denaturation and incubation for 18 hours at 25 °C (Oxford GlycoSystems).

Circular dichroism (CD) measurements

Native *Ber e 1* protein from Brazil nut seeds was prepared essentially as described.¹³ CD measurements were performed with a Jasco J-810 spectropolarimeter in the far UV region using a protein concentration of 0.03–0.32 mg/ml and a 0.1 cm cell pathlength. The buffers used were 10 mM phosphate buffer (pH 6.8) and 0.1 M glycine buffer (pH 2.2). The instrument was calibrated with d-10-camphorsulphonic acid. Spectra were acquired using a scan speed of 20 nm/minute, with a 2 mm band-

width and a two seconds integration time. The spectra were measured 20 times, averaged, and corrected by subtraction of the solvent spectrum obtained under identical conditions. Native and recombinant *Ber e 1* proteins, as well as recombinant SFA8 were examined by CD.

Heat-induced unfolding

Heat-induced unfolding transitions were monitored at fixed wavelength of 222 nm using the same protein and buffer concentration as above. The temperature was increased monotonically from 25 °C to 95 °C at a rate of 0.5 deg.C/minute. A spectrum was recorded at 95 °C, and the reversibility of the phenomenon was assayed by measuring the spectra of the sample cooled down in a single step to 25 °C. Data were acquired with a reading frequency of 1/60 second, eight seconds integration time and 2 nm bandwidth. The resulting curves were corrected for the contribution of the solvent obtained under identical heating conditions. The protein concentration was determined using the BCA protein assay reagent kit (Pierce, Illinois).

Chemical-induced unfolding

Samples of the protein were incubated overnight at 25 °C in the presence of various concentrations in guanidinium chloride (i.e. 0, 1.5, 3, 4.5 and 6 M). The denaturant concentration was determined from refractive index measurements using a R5000 hand refractometer from Atago. Spectra were recorded using the same parameters as for heat-induced unfolding and were corrected for the contribution of the solvent obtained under identical heating conditions.

Digestion of *Ber e 1* and HPLC analysis

Non-reduced protein (200–500 μg) was digested overnight at 37 °C with (5 μg) sequencing grade-modified trypsin (Promega) in (100 μl) ammonium hydrogen carbonate buffer (150 mM, pH 8.1). Reverse phase HPLC was carried out on a model 130A (Applied Biosystems) equipped with a Phenomenex Jupiter C18 column (2.1 mm \times 250 mm). Aqueous formic acid (0.1% (v/v) solvent A) and 70% (v/v) acetonitrile in solvent A (solvent B) were used in a linear gradient of 0–100% B over 60 minutes following an initial period of five minutes at 0% B. The flow rate was 150 μl /minute and the peptides detected at 220 nm with manual collection. Whenever required, the large peptide fragments collected from HPLC were reduced with DTT (20 μl , 100 mM) and incubated at room temperature for one hour. Cyanogen bromide cleavage was carried out with large crystals in 70% (v/v) formic acid and incubated overnight at room temperature in the dark. The resulting fragments were re-run on HPLC as detailed above and peaks collected for subsequent mass and N-terminal sequencing analysis.

Mass spectrometry

Masses of glycosylated proteins were derived by analysis using a MALDI-TOF (matrix-assisted laser desorption ionisation-time of flight) spectrometer on a Voyager-DE™ Biospectrometry™ Workstation (Applied Biosystems, USA) operating at 20,000 V accelerating potential and in delayed extraction mode. The sample was prepared at 2 pmol/ μl in water and mixed with

saturated sinipinic acid solution (in 50% acetonitrile, 0.1% TFA).

Masses of intact peptide complexes and individual peptide fragments were derived by analysis using a Micromass QTOF-2 hybrid (quadrupole time of flight mass spectrometer) utilising MassLynx software. Peptide solutions were introduced into the sample cone as a fine spray from the end of a thin-walled glass capillary using, typically, a capillary voltage of 800 V. The masses of large peptide complexes and other multiply charged ion species were deconvoluted to singly charged masses using the maximum entropy MaxEnt3 software (Micromass). Peptides were selected for fragmentation in MS-MS mode in the mass spectrometer and subjected to collision energies ranging from 20–80 eV over a period of time. All data collected from each fragmentation experiment were deconvoluted using MaxEnt3 software. *De novo* sequence assignments were carried out interactively utilising the peptide-sequencing portion of the BioLynx menu included in the MassLynx software package.

Amino acid sequence analysis

N-terminal sequencing of proteins was carried out using blotted PVDF membranes or eluted HPLC peaks onto Bolybrene-coated glass fibre discs and sequenced using an automated protein sequencer Model 473A (Applied Biosystems) equipped with an on-line PTH Chromatograph and model 610A software (Applied Biosystems).

Homology modelling of Ber e 1 protein

On the basis of the results of 3D-PSSM⁴² comparisons of the Ber e 1 amino acid sequence, the experimentally determined disulphide connectivity, and a structure-based sequence alignment of eight members of the 2S albumin "four helices; folded leaf; right-handed superhelix; disulphide-rich" fold family produced previously,³⁷ we used the structure of the napin seed storage protein (PDB code 1PNB³⁹) as a structural template. The appropriate sections of the Ber e 1 amino acid sequence were directly substituted onto the 1PNB structure according to the alignment in Figure 7(a). Side-chain clashes in the resulting model were relieved by altering torsion angles by inspection. The structure was then covered in a 5 Å layer of water and energy minimised until the absolute derivative of co-ordinates with respect to the energy fell below 0.01 kcal mol⁻¹ Å⁻¹. Analysis of the structure in the program Procheck⁴⁰ demonstrated that the minimised model had acceptable geometry, as shown by the Ramachandran angles and G factors (data not shown). Visualisation, building and manipulation of the structures was carried out using InsightII 2.35, and energy calculations were performed using Discover 2.95 (MSI) using the consistent valence force field.

Atomic co-ordinates

The model structure has been deposited in the PDB database under the accession number 1GYS.

Acknowledgements

This work was sponsored by the UK Food Standards Agency (T07019) at the University of Nottingham.

References

1. Spies, J. R., Coulson, E. J., Chambers, D. C., Bernton, H. S., Stevens, H. & Shimp, J. H. (1951). The chemistry of allergens. XI. Properties and composition of natural proteoses isolated from oilseeds and nuts by the CS-1A procedure. *J. Am. Chem. Soc.* **73**, 3995–4001.
2. Spies, J. R. (1974). Allergens. *J. Agric. Food Chem.* **22**, 30–36.
3. Atkins, F. M., Wilson, M. & Bock, S. A. (1988). Cottonseed hypersensitivity: new concerns over an old problem. *J. Allergy Clin. Immunol.* **82**, 243–250.
4. Youle, R. & Huang, A. H. C. (1981). Occurrence of low molecular weight and high cysteine containing albumin storage protein in oilseeds of disperse species. *Am. J. Bot.* **68**, 44–48.
5. Pastorello, E. A., Pompei, C., Pravettoni, V., Brenna, O., Farioli, L., Trambaioli, C. & Conti, A. (2001). Lipid transfer proteins and 2S albumins as allergens. *Allergy*, **56**, 45–47.
6. Takizawa, T., Arakawa, H., Tokuyama, K. & Morikawa, A. (2001). Identification of allergen fractions of wheat flour responsible for anaphylactic reactions to wheat products in infants and young children. *Int. Arch. Allergy Immunol.* **125**, 51–56.
7. Sandiford, C. P., Tatham, A. S., Fido, R., Welch, J. A., Jones, M. G., Tee, R. D. *et al.* (1997). Identification of the major water/salt insoluble wheat proteins involved in cereal hypersensitivity. *Clin. Expt. Allergy*, **27**, 1120–1129.
8. Shewry, P. R. & Pandya, M. J. (1999). The 2S albumin storage proteins. *Seed Proteins*, Kluwer Academic Publishers, The Netherlands pp. 619–643.
9. Saalbach, G., Rosso, M. & Schumann, U. (1996). The vacuolar targeting signal of the 2S albumin from brazil nut resides at the C terminus and involves the C-terminal propeptide as an essential element. *Plant Physiol.* **112**, 975–985.
10. Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., van Damme, J. *et al.* (1988). Determination of the processing sites of an Arabidopsis 2S albumin and characterisation of the complete gene family. *Plant Physiol.* **87**, 859–866.
11. Ampe, C., van Damme, J., Castro, L. A. B., Sampaio, M. J. A. M., van Montagu, M. & Vandekerckhove, J. (1986). The amino acid sequence of the 2S sulphur-rich proteins from seeds of brazil nut (*Bertholletia excelsa* H.B.K.). *Eur. J. Biochem.* **159**, 597–604.
12. Castro, L. A. B., Lacerda, Z., Aramayo, R. A., Sampaio, M. J. A. M. & Gander, E. S. (1987). Evidence for a precursor molecule of brazil nut 2S seed proteins from biosynthesis and cDNA analysis. *Mol. Gen. Genet.* **206**, 338–343.
13. Sun, S. S. M., Altenbach, S. B. & Leung, F. W. (1987). Properties, biosynthesis and processing of a sulfur-rich protein in brazil nut (*Bertholletia excelsa* H.B.K.). *Eur. J. Biochem.* **162**, 477–483.
14. Kortt, A. A. & Caldwell, J. B. (1990). Low molecular weight albumins from sunflower seed: identification of a methionine-rich albumin. *Phytochemistry*, **29**, 2805–2810.

15. Kortt, A. A., Caldwell, J. B., Lilley, G. G. & Higgins, T. J. V. (1991). Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). *Eur. J. Biochem.* **195**, 329–334.
16. Pastorello, E. A., Farioli, L., Pravettoni, V., Spano, M., Conti, A., Ansaloni, R. *et al.* (1998). Sensitization to the major allergen of brazil nut is correlated with clinical expression of allergy. *J. Allergy Clin. Immunol.* **102**, 1021–1027.
17. Nordelee, J. A., Taylor, S. L., Townsend, J. A., Thomas, L. A. & Bush, R. K. (1996). Identification of brazil nut allergen in transgenic soybeans. *N Engl. J. Med.* **334**, 688–692.
18. Kelly, J. D., Hlywka, J. J. & Hefle, S. L. (2000). Identification of sunflower seed IgE-binding proteins. *Int. Arch. Allergy Immunol.* **121**, 19–24.
19. Schwartz, R. H. (1985). T lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Annu. Rev. Immunol.* **3**, 237–255.
20. Robotham, J. M., Teuber, S. S., Sathe, S. K. & Roux, K. H. (2002). Linear IgE epitope mapping of the English walnut (*Juglans regia*) major food allergen, Jug r 1. *J. Allergy Clin. Immunol.* **109**, 143–149.
21. Ommen, A. M., Kelly, J. K., Benson, A. K., Hefle, S. L. & Taylor, S. L. (2000). Characterization of 2S brazil nut IgE-binding epitopes. *J. Allergy Clin. Immunol.* **103**, pS101–pS134.
22. van Regenmortel, M. H. V. (1989). Structural and functional approaches to the study of protein antigenicity. *Immunol. Today*, **10**, 266–272.
23. Ferreira, F., Ebner, C., Kramer, B., Casari, G., Briza, P., Kungl, A. J. *et al.* (1998). Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. *FASEB J.* **12**, 231–242.
24. Spencer, J. A., Jeenes, D. J., MacKenzie, D. A., Haynie, D. T. & Archer, D. B. (1998). Determinants of the fidelity of processing glucoamylase–lysozyme fusions by *Aspergillus niger*. *Eur. J. Biochem.* **258**, 107–112.
25. MacKenzie, D. A., Kraunsoe, J. A. E., Chesshyre, J. A., Lowe, G., Komiyama, T., Fuller, R. S. & Archer, D. B. (1998). Aberrant processing of wild-type and mutant bovine pancreatic trypsin inhibitor secreted by *Aspergillus niger*. *J. Biotechnol.* **63**, 137–146.
26. Parekh, R., Forrester, K. & Wittrup, D. (1995). Multi-copy overexpression of bovine pancreatic trypsin inhibitor saturates the protein folding and secretory capacity of *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **6**, 537–545.
27. Ledgerwood, E. C., George, P. M., Peach, R. J. & Brennan, S. O. (1995). Endoproteolytic processing of recombinant proalbumin variant by the yeast Kex2 protease. *Biochem. J.* **308**, 321–325.
28. Julius, D., Schekman, R. & Thorner, J. (1984). Glycosylation and processing of prepro- α -factor through the yeast secretory pathway. *Cell*, **36**, 309–318.
29. Takahashi, K., Takai, T., Yasuhara, T., Yokota, T. & Okumura, Y. (2001). Effects of site-directed mutagenesis in the cysteine residues and the N-glycosylation motif in recombinant Der f 1 on secretion and protease activity. *Int. Arch. Allergy Immunol.* **124**, 454–460.
30. Best, E. A., Stedman, K. E., Bozic, C. M., Hunter, S. W., Vailes, L., Chapman, M. D. & McDermott, M. J. (2000). A recombinant group 1 house dust mite allergen, rDer f 1 with biological activities similar to those of the native allergen. *Protein Expr. Purif.* **20**, 462–471.
31. Bretthauer, R. K. & Castellino, F. J. (1999). Glycosylation of *Pichia pastoris*-derived proteins. *Biotechnol. Appl. Biochem.* **30**, 193–200.
32. Gemmill, T. R. & Trimble, R. B. (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta*, **1426**, 227–237.
33. Duman, J. G., Miele, R. G., Liang, H., Grella, D. K., Sim, K. L., Castellino, F. J. *et al.* (1998). O-Mannosylation of *Pichia pastoris* cellular and recombinant proteins. *Appl. Biochem.* **28**, 39–45.
34. van Oort, E., Heer, P. G., Lerouge, P., Faye, L., Aalberse, R. C. & van Ree, R. (2001). Immunochemical characterisation of two *Pichia pastoris*-derived recombinant group 5 *Dactylis glomerata* isoallergens. *Int. Arch. Allergy Immunol.* **126**, 196–205.
35. Vailes, L. D., Kinter, M. T., Arruda, L. K. & Chapman, M. D. (1998). High level expression of cockroach allergen, Bla g 4 in *Pichia pastoris*. *J. Allergy Clin. Immunol.* **101**, 274–280.
36. Letourneur, O., Gervasi, G., Gaia, S., Pages, J., Watelet, B. & Jolivet, M. (2001). Characterization of *Toxoplasma gondii* surface antigen I (SAGI) secreted from *Pichia pastoris*: evidence of hyper-glycosylation. *Biotechnol. Appl. Biochem.* **33**, 35–45.
37. Pandya, M. J., Sessions, R. B., Williams, P. B., Dempsey, C. E., Tatham, A. R., Shewry, P. R. & Clarke, A. R. (2000). Structural characterisation of a methionine-rich, emulsifying protein from sunflower seed. *Proteins: Struct. Funct. Genet.* **38**, 341–349.
38. Egorov, T. A., Odintsova, T. I., Musolyamov, A. Kh., Fido, R., Tatham, A. S. & Shewry, P. R. (1996). Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily. *FEBS Letters*, **396**, 285–288.
39. Rico, M., Bruix, M., González, C., Monsalve, R. I. & Rodríguez, R. (1996). ¹H NMR assignment and global fold of napin *BnIb*, a representative 2S albumin seed protein. *Biochemistry*, **35**, 15672–15682.
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
41. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (2001). *Current Protocols in Molecular Biology*, Wiley, New York.
42. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 501–522.

Edited by A. R. Fersht

(Received 3 July 2002; received in revised form 20 September 2002; accepted 27 September 2002)