

The expression and processing of two recombinant 2S albumins from soybean (*Glycine max*) in the yeast *Pichia pastoris*

Jing Lin^a, Roger Fido^b, Peter Shewry^b, David B. Archer^a, Marcos J.C. Alcocer^{a,*}

^a*School of Biology, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

^b*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

Received 2 October 2003; received in revised form 1 December 2003; accepted 1 December 2003

Available online 17 December 2003

Abstract

Soybean seeds contain two 2S albumin storage proteins (AL1 and AL3) which may contribute to their industrial processing quality and allergenicity. We show that these proteins (AL1 and AL3) are well expressed by the methylotrophic yeast *Pichia pastoris* and that one of the secreted proteins (AL3) has a similar conformation and stability to that purified from soybean seeds. Further, we show that the subunits are post-translationally processed within the same loop region as the native protein but with some differences in the precise sites. This internal processing provides useful information on the endoproteolytic activity in *P. pastoris*. We also show that, similar to many plant allergens, the 2S albumins from soybean are stable to heat and chemical treatments.

© 2004 Elsevier B.V. All rights reserved.

Keywords: 2S albumin; Soybean protein; Plant allergen; *Pichia pastoris*; Proteolytic processing

1. Introduction

Soybean is a major source of protein for human and animal nutrition in many countries. Like other legumes, the nutritional quality of soybean is limited by low contents of the sulfur-containing amino acids, cysteine and methionine. Several strategies have been adopted to improve the nutritional value of soybean [1,2], including over-expression of an exogenous gene encoding a methionine-rich protein [3,4]. More recently, the over-expression of endogenous rather than exogenous methionine-rich proteins has been suggested as an alternative strategy aiming at a more stable expression with less impact on the normal biology of the seed [5].

The 2S albumins are a group of storage proteins that occur widely in seeds of dicotyledonous plants [6,7]. In several species they include specific components that are rich in methionine, notably in Brazil nut and sunflower, and they are consequently candidates for nutritional supplementation. The clinical implication of such use, particularly in relation to the intrinsic allergenicity of the 2S albumin family, is a cause for concern. Historically, the 2S albumins were identified as

allergens in occupational diseases in the 1920s but were only fully characterized as the major allergens in the late 1970s [6]. More recently, the well-publicised case of the allergenic Brazil nut 2S albumin (*Ber e 1*) being introduced into transgenic legumes highlighted the potential problems [8].

So far, two 2S albumins have been identified from soybeans: 2S albumin 1 (AL1, GenBank U71194) and 2S albumin 3 (AL3, GenBank U71195). Despite their low sequence similarity with other members of the family, these 2S albumins possess the characteristic cysteine skeleton of the class. The three-dimensional structures determined for 2S albumins from oilseed rape, castor bean and sunflower show a conserved antiparallel bundle of four α -helices held together by four disulfide bonds in a right-handed fold and a similar structure is expected for the soybean proteins [9–11]. The 2S albumins so far identified from soybean differ from other albumins in having an unusually high content of charged amino acids such as aspartic acid (AL1, 5.66%; AL3, 12.28%) and glutamic acid (AL1, 16.99%; AL3, 15.38%). Although the main biological role for 2S albumins is thought to be storage, the demonstration of antimetabolic activity in human cells studies could indicate a regulatory role during early stages of seed development [12]. This antimetabolic activity in mammalian cells is claimed to have

* Corresponding author. Fax: +44-115-951-3251.

E-mail address: Marcos.Alcocer@Nottingham.ac.uk (M.J.C. Alcocer).

contributed to the low incidence of colon cancer in some Asian countries where soybean is a staple diet [13,14].

Allergic reactions to soybean are prevalent in several regions of the world where the consumption is high. Similar to many food allergens, soybean allergy is more prevalent in infants with symptoms ranging from mild skin, gastrointestinal and respiratory reactions [15,16], to severe systemic reactions including anaphylaxis [17]. The allergenicity of soybean is known to reside in the protein fraction [18] and several major allergens covering a large number of protein families have been isolated and characterized, including food allergens: Kunitz trypsin inhibitors [17], α -subunit of the 7S vicilin-like β -conglycinin (Gly m Bd 60K) [19], an oil-body-associated protein (Gly m Bd 30K) which is related to cysteine proteases [20], a vicilin-like glycoprotein (Gly m Bd 28K) [21], the 11S protein glycinin [22], profilin (Gly m 3) [23], a stress-induced protein (Gly m 4) [24], and respiratory allergens: soybean hydrophobic proteins (Gly m 1.0101, Gly m 1.0102) [25], a hull protein (Gly m 2) [26]. Interestingly, the 2S albumins from soybean have not yet been reported as major allergens.

In order to investigate the structure–function relationships of the 2S albumins from soybean, we report their expression, purification and biochemical characterization. We show that these proteins are well expressed by the methylotrophic yeast *Pichia pastoris* and that one of the secreted proteins has a similar conformation and stability to that purified from soybean seeds. Furthermore, we show that the subunits are post-translationally processed within the same loop region as the native protein but with some differences in the precise sites. We also show that, similarly to other well-characterized plant allergens, the 2S albumins from soybean are stable to heat and chemical treatments.

2. Materials and methods

2.1. Chemicals

Unless otherwise stated, all the chemicals used were from Sigma (Dorset, UK). Immobilon polyvinylidene fluoride (PVDF) membrane (pore size, 0.45 μ m) was from Millipore (Bedford, MA, USA), restriction enzymes and T₄ DNA ligase from New England Biolabs (Beverly, MA, USA), Taq DNA polymerase from Abgene (Surrey, UK) and DNA purification kits from Qiagen (West Sussex, UK).

2.2. DNA sequences

Two plasmids containing the cDNA encoding two soybean 2S albumins, AL1 and AL3, respectively, were kindly provided by Pioneer Hi-Bred International, Inc., Des Moines, IA, USA. AL1 and AL3 have been filed with GenBank with the accession numbers U71194 and U71195, respectively.

2.3. Cell lines

Plasmid pPIC9 and *P. pastoris* strain GS115 were from Invitrogen (Carlsbad, CA, USA). *Escherichia coli* XL1Blue (Stratagene, La Jolla, CA, USA) was employed in most transformation procedures according to standard protocols [27].

2.4. Antibodies

The antibody used to detect albumins in dot-blot analysis was raised against native 2S albumin (nAL) in rabbits and provided by Pioneer Hi-Bred International. The antisera used in the enzyme-linked immunosorbent assay (ELISA) were raised in rats against the purified recombinant soybean albumins at the University of Nottingham.

2.5. Culture media

Media were prepared according to Invitrogen's Instruction Manual (A Manual of Methods for Expression of Recombinant Proteins in *Pichia pastoris*). *P. pastoris* strain GS115 was maintained in yeast extract peptone dextrose (YPD) broth. Regeneration dextrose (RD) agar without histidine was used for plasmid selection. Buffered minimal glycerol media (BMG) broth and buffered minimal methanol medium (BMM) broth were used for production of the recombinant proteins.

2.6. Construction of the recombinant albumin (rAL) expression plasmid for *P. pastoris*

To secrete rAL from *P. pastoris*, two expression plasmids, each carrying one soybean albumin cDNA sequence, were constructed as follows. The 2S albumin genes were amplified by PCR using the primers P1: 5'-AGAAGGGG-TATCTCTCGAGAAAAGAGAGGCTGAAGCTTC-CCAAATGGCAGCACCAGCA-3'/P2: 5'-CCTAGGGAATTCTCAGTCATCGGAGGACAAGTCGC-3' and P1/P3: 5'-AGGGAATTCTTAGTCATCGGAGGA-ACAAGTCGCA-3' for AL1 and AL3, respectively. *Xho*I and *Eco*RI restriction sites were incorporated and are shown in italics. The PCR products were digested with *Xho*I and *Eco*RI, and ligated into pPIC9 previously digested with the same enzymes. The resulting plasmids were transformed into *E. coli* XL1Blue and selected on ampicillin (75 μ g/ml). Plasmids containing the inserted gene were purified and DNA sequence analysis was performed using the Sanger method with dye terminators (Applied Biosystems, Warrington, UK).

2.7. Construction of rAL-producing *P. pastoris*

The expression plasmids containing the 2S albumin genes were linearized by overnight digestion with *Sal*I at 37 °C for efficient integration into the *his4* locus of the *P.*

pastoris genome. GS115 cells were made competent and transformed with the linearized plasmids by electroporation (Micropulser, Biorad, Hertfordshire, UK) according to the manufacturer's instructions. Transformant colonies (RD agar without histidine), grown at 30 °C for 3–4 days, were assessed by dot-blot analysis as described previously [28], except that the first antibody used (diluted 1:2000) was raised against the native soybean 2S albumin in rabbit. The best producers selected were kept for further analysis.

2.8. Large-scale production of rAL

The selected rAL-producing strains were grown in five 4-l flasks each containing 1.2 l of BMG medium at 30 °C until they reached high cell density (1.7×10^8 – 6.8×10^8 cells/ml). The cultures were then centrifuged at $7000 \times g$ for 10 min. The supernatant was discarded and the same amount of BMM medium was added to resuspend the cells. Incubation was continued for up to 80 h and methanol (0.5% v/v) was added every 24 h to compensate for evaporation. Cell growth was monitored at 600 nm and the expression level of the secreted rAL was assessed by ELISA inhibition assay.

Following incubation, the culture supernatant was separated from the cells by centrifugation at $7000 \times g$ for 10 min and filtered through a 0.45 µm filter. The volume of the supernatant was then reduced to 100 ml and the salt eliminated by buffer exchange with 2 l 10 mM citric acid buffer (pH 4.3) using Viva Flow 200 (5000 MWCO) (Vivascience, Lincoln, UK) according to the manufacturer's instructions. This concentrated supernatant was stored at –20 °C until use.

2.9. Protein determination

Concentrations of total proteins and the purified rAL were determined using the bicinchoninic acid (BCA) protein assay reagent kit (Pierce, Rockford, IL, USA) with bovine serum albumin (BSA) as standard. 2S albumin concentration was determined by ELISA inhibition assay as follows. A 96-well microtiter plate (Dynatech Microfluor, VA, USA) was coated (100 µl/well) overnight at 4 °C with 1 µg/ml purified rAL. After washing three times with PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 , 0.24 g KH_2PO_4 per litre, pH 7.4), 0.05% v/v Tween 20 (PBST), unoccupied binding sites were blocked for 1 h at room temperature with BSA (1% w/v) in PBST. The plate was then incubated for 2 h at room temperature with the antisera (100 µl/well), diluted 1:5000 in PBST plus 1% w/v BSA and 20% v/v *P. pastoris* supernatant extract, which did not contain the target protein. The antisera used were raised against the purified recombinant soybean albumins, and previously incubated for 1 h with different dilutions (0.2–625 µg/ml) of protein samples as inhibitors. Plates were washed three times with PBS, followed by

incubation for 1 h at room temperature with alkaline phosphatase-conjugated anti-rat IgG (Sigma) at 1:1000 dilution in PBS. After washing three times with PBS, the plates were then incubated with Sigma FAST pNPP (*p*-nitrophenyl phosphate) substrate (100 µl/well) for 30 min. Optical densities were measured at 450 nm. Concentrations of rAL were determined according to the standard curve obtained by using various concentrations of the purified rAL as inhibitors in the ELISAs.

2.10. Purification of rAL

Soybean albumin was separated from the other proteins secreted by *P. pastoris* using Biorad Duoflow system equipped with three in-series 5 ml Hi-Trap heparin columns (Amersham-Pharmacia, Uppsala, Sweden). The concentrated supernatant was loaded onto the heparin column, previously equilibrated with 10 mM citric acid buffer (pH 4.3) and eluted using a step gradient of NaCl (0.2–1.0 M) at a flow rate of 3 ml/min. The column effluent was monitored at 280 nm. The collected fractions containing rAL were dialyzed against distilled water for 24 h at 4 °C. The rAL was then further purified by reversed-phase HPLC on a C_8 column (10 cm $\times$ 10 mm, Varian, Walnut Creek, CA, USA), in 0.1% formic acid in water and eluted with a linear gradient of 0–100%, 100% acetonitrile in 0.1% formic acid at a flow rate of 2.5 ml/min for 25 min. The eluted peak was lyophilised and stored at –20 °C.

2.11. Purification of the native 2S albumin from soybean seed

The native soybean 2S albumin was purified from soybean seed based on a method developed by Pioneer Hi-Bred International. Soybean seeds were ground into a fine powder, defatted by extraction with hexane (1:4 w/v) and dried in vacuum. Soluble proteins were extracted from 100 g of defatted soybean powder with 400 ml albumin extraction buffer (10% v/v DMSO, 0.5% v/v butan-1-ol, 100 mM KCl, 83 mM sodium acetate, pH 5.2). Following centrifugation at $6000 \times g$ for 15 min, the supernatant was dialyzed against 0.5% v/v butan-1-ol, 100 mM KCl, 83 mM sodium acetate buffer, pH 5.2, lyophilised, and dissolved in distilled water (250 mg/ml). Insoluble proteins were removed by centrifugation at $6000 \times g$ for 15 min and the supernatant was loaded onto a Sephacryl S-100 column (2 cm $\times$ 55 cm, Pharmacia) run (0.5 ml/min) with the albumin extraction buffer. Fractions containing the 2S albumins (~14 kDa) were dialyzed against distilled water, lyophilised, dissolved in 20 mM Tris/HCl, pH 8.5 (20 mg/ml) and further fractionated by ion-exchange chromatography using the BioCAD perfusion chromatography system equipped with a POROS 20 HQ (4.6 mm $\times$ 10 cm) column, eluted with a gradient of NaCl (0–750 mM) in 20 mM Tris/HCl, pH 8.5 at a flow rate of 5 ml/min. The peak, which was

eluted at 360 mM NaCl and contained the 2S albumin, was then further purified by reversed-phase HPLC on a C₁₈ column (25 cm × 10 mm, Phenomenex), in 0.07% trifluoroacetic acid (TFA) in water and eluted with a linear gradient of 0–50%, 100% acetonitrile in 0.05% TFA at a flow rate of 2.5 ml/min for 40 min and the purified 2S albumin eluted after 25 min. The eluted peak was lyophilised and stored at –20 °C.

2.12. SDS-PAGE, N-terminal sequencing and mass spectrometry

SDS-PAGE was performed under reducing conditions using 12% Bis-Tris NuPAGE gel (Invitrogen) according to the manufacturer's instructions. The protein bands were either stained with Coomassie Brilliant Blue R-250, or blotted onto PVDF membrane for N-terminal sequencing using NOVEX Xcell II blotting apparatus (Invitrogen) according to the manufacturer's instructions. Determination of N-terminal sequences was performed on an Applied Biosystems model 473A automated protein sequencer. Masses of proteins were determined under reducing conditions using a Micromass MALDI-TOF MS.

2.13. Circular dichroism (CD) measurements

CD measurements of the native and recombinant soybean 2S albumins were performed with a JASCO J-810 spectropolarimeter in the far UV region using a protein concentration of 0.1–0.2 mg/ml (6.9×10^{-6} – 1.38×10^{-5} mol/l) and a 0.1 cm cell pathlength. The buffers used were 10 mM phosphate buffer (pH 6.8) and 10 mM glycine buffer (pH 2.2). Both heat-induced unfolding and chemical-induced unfolding were measured using guanidinium chloride as described previously [28].

3. Results

3.1. Construction of rAL-producing *P. pastoris* strains

Two pPIC9-based plasmids, each containing the cDNA sequence encoding a mature soybean 2S albumin, were constructed for the expression of rAL in *P. pastoris*. In order to direct the expressed protein into the secretion pathway, the cDNA sequence was fused with the methanol-inducible 5'-AOX1 promoter, the sequence encoding the α -factor secretion signal, and the 3'-AOX transcriptional terminator. As shown in Fig. 2, the amino acids that constitute the Kex2 processing site (KR) and Ste13 putative processing site (EAEA) were included to facilitate proteolytic processing during secretion.

Following electroporation, the linearized expression vectors were integrated into the *his4* locus of the chromosome and selected for histidine auxotrophic mutants. Cell lines

expressing high levels of recombinant 2S albumins were identified by immuno-dot-blot screening.

3.2. Secretion of rAL by *P. pastoris* and purification of rAL

Expression of the albumin-encoding gene was induced in medium containing methanol and the expressed rAL secreted into the culture supernatants. In order to facilitate protein purification, only minimal medium was used. After 80 h of induction, the expression levels reached 4.4 mg/l for rAL1 and 4.7 mg/l for rAL3, representing 21.5% and 23.4% of the total secreted proteins in the culture supernatants, respectively.

After concentration and buffer exchange using a tangential flow system, the concentrations of 2S albumins in culture supernatant were from ~20% to ~75% of the total proteins using Heparin columns, as reported for other members of the 2S albumin family [28]. In order to separate the lipid-bound fraction from the recombinant proteins, a further purification step using reversed-phase HPLC was carried out. The best-expressing rAL1 and rAL3 cell lines yielded 1.8 and 2.1 mg of the pure albumin per litre of culture, respectively.

3.3. Purification of native soybean 2S albumin

The nAL was purified from soybean seed using a combination of three types of chromatography: size exclusion, ion-exchange, and reversed-phase HPLC. nAL (1.8 mg) was purified per 10 g of soybean seed. The purified nAL appeared as two bands with apparent molecular weights of 9 and 5 kDa when separated by SDS-PAGE under reducing conditions (Fig. 1). N-terminal sequence analyses were performed and the results are shown in Fig. 2. Although many varieties of soybean exist and intra-species polymorphic variation may be expected, the N-terminal sequence of the large subunit (EGKDEDEEEE) was identical to the GenBank sequence of AL3 (U71195). Mass

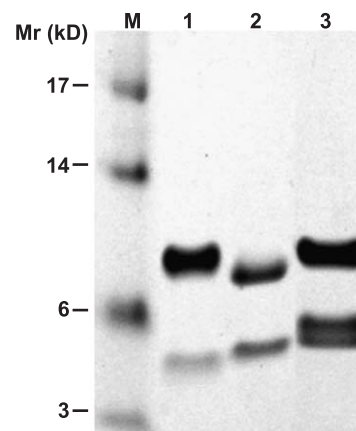


Fig. 1. SDS-PAGE of the purified native and recombinant soybean 2S albumins under reducing conditions in 12% Bis-Tris NuPAGE gel (Invitrogen). Lane M, molecular weight marker (See Blue Plus2 pre-stained protein standard); lane 1, nAL3; lane 2, rAL1; lane 3, rAL3.

```

10-----1-----11-----21-----31-----41-----51-----61-----71-----
rAL1: α-mating...siaakeegvslekreaea▼SKWQHQQESCREQLKGINLNPCEHIMEKIQAGRGEDGSDDEDHIL↑IR↑TMPGRINYI▼R▼KKEGKKEEBEGHMQ
jpred: -----HHHHHH-----HHHHHHHH-----EEE EE-----EEE E-----HH
SUBsec: .....HHHHHHHH.....HHHHHHHH.....LLLL.....L.....HHHHHH
SUBacc: .e....e.bie.....e.b..be.....e...... ....e.....
-----81-----91-----101-----111-----121-----131-----
rAL1: CCTEMSELKSPICQCKALQKIMDNQSEQLGKQKMERELMNLAIRCLGPMIGCDLSSDD
jpred: HH-----HHHHHHHHHHHH-----HHHHHHHHHH-----
SUBsec: HHHHH...LL.....HHHHHHHH.....HHHHHHHH.....LL.....LL.LL
SUBacc: bbeib.eie.....b.e.be.....e.....ei.e.....ee
10-----1-----11-----21-----31-----41-----51-----61-----71-----
rAL3: α-mating...siaakeegvslekreaea▼SKWQHQQDSRCRQLQGVNLTPTCEKHIMEKIQGRGDDDDDDDDN HIL↑RTM↑RGRINY↑IRR▼NEGKDEDEEBEGHMQ
nAL3: mtkftillislfciahtcsa↑SKWQHQQDSRCRQLQGVNLTPTCEKHIMEKIQGRGDDDDDDDDN HIL RTM RGRINY IRRN▼EGKDEDEEBEGHMQ
jpred: -----HHHH-----HHHHHHHH-----H HHH H-----EEE-----HH
SUBsec: .....HHHHHHHH..LLL..HHHHHHHH..LLLL......LLL.....L LLLLLL...HHHH
SUBacc: ee.e...e.bee.bee.e....e.b.e.b....eee.e......eeee.....e
-----81-----91-----101-----111-----121-----131-----
rAL3: CCTEMSELRS PKCQCKALQKIMENQSEELBEKQKKMEKELINLATMCRFGPMIQCDLSSDD
nAL3: CCTEMSELRS PKCQCKALQKIMENQSEELBEKQKKMEKELINLATMCRFGPMIQCDLSSDD
jpred: HHHHH-----HHHHHHHHHHHH-----HHHHHHHHHH-----
SUBsec: HHHHHHH...L.....HHHHHHHH.....HHHHHHHH.....LLLL.....LLLLL
SUBacc: bbeib.e.....bb.i.e.....e.e.....ee.Iee.....eee

```

spectrometry analysis under reducing conditions showed that the masses of the two subunits were 8975.14 and 5138.30 (data not shown), which is an exact match with

the predicted masses of the subunits derived from the gene sequence (U71195). Therefore, the purified native soybean 2S albumin is confirmed to be the native AL3 (nAL3).

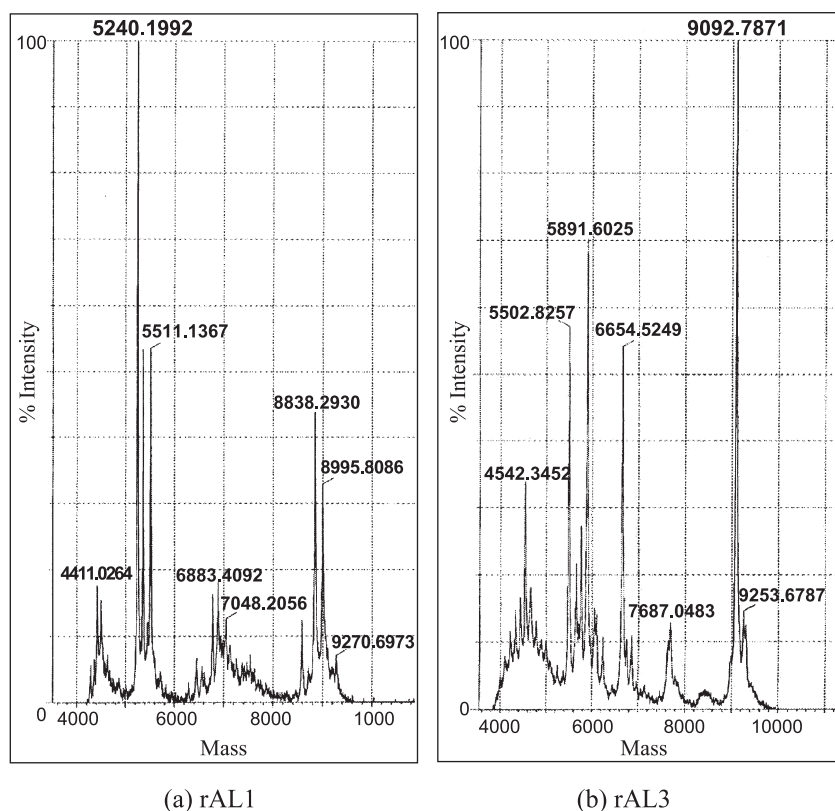


Fig. 3. MALDI-TOF mass spectra of the recombinant soybean albumins under reducing conditions.

3.4. rAL processing

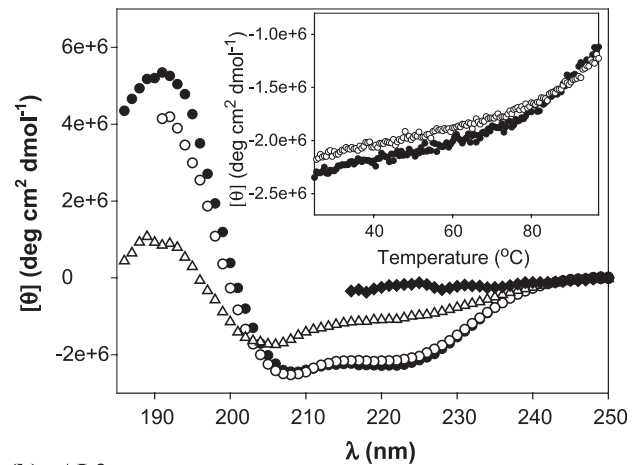
Purified rAL1 gave only two bands corresponding to the large and small subunits when separated by SDS-PAGE under reducing conditions (Fig. 1, lane 2). However, N-terminal amino acid sequencing showed the presence of two forms of the large subunit differing only in a single Arg residue (sequences RKKEGKEEEEEEGHMQ/KKEGKEEEEEEGHMQK) (Fig. 2), whereas mass spectrometry (Fig. 3) showed two forms of each subunit (masses 5240.20 and 5511.14 Da, 8838.29 and 8995.80 Da, respectively). Similarly, SDS-PAGE of rAL3 showed two small subunit bands (Fig. 1, lane 3), although neither of these bands corresponded exactly in mobility to that in the nAL3 (Fig. 1, lane 1). In this case, N-terminal sequencing showed only single sequences for each subunit (Fig. 2), while mass spectrometry (Fig. 3) showed one large subunit (mass 9092.79 Da) and three small subunits (masses 5502.83, 5891.60, and 6654.52 Da). These results indicate that both Kex2 (KR) and Ste13 (EAEA) processing sites were recognized, resulting in complete removal of the α -mating type leader sequence. The possible processing events occurring in the secretory pathway of *P. pastoris* are summarized in Table 1.

Table 1
Proteolytic processing sites in 2S albumins expressed by *P. pastoris* and their putative proteases

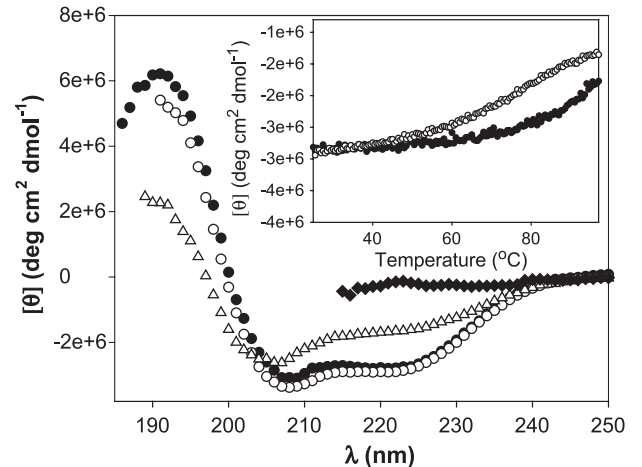
Sequences	Kex1	Kex2/ Yapsin family endoprotease	Ste13	Protein	Reference
EEGVSLEKR [▼] <i>ea</i> _Δ SKWQ		+	+	rAL1	this work
SDEDHIL [□] <i>ir</i> [▼] TMPGRINYI	+	+		rAL1	this work
TMPGRINYI [▼] RKKEGKEEEE		?		rAL1	this work
MPGRINYIR [▼] KKEGKEEEE		+		rAL1	this work
EEGVSLEKR [▼] <i>ea</i> _Δ SKWQH		+	+	rAL3	this work
DDDDNHIL [□] <i>r</i> [▼] TMRGRINYI	+	+		rAL3	this work
DNHILRTM [□] <i>r</i> [▼] GRINYIRRN	+	+		rAL3	this work
RGRINY [□] <i>irr</i> [▼] NEGKDEDEE	+	+		rAL3	this work
EEGVSLEKR [▼] EAEAEQEECR		+		Ber e 1	[28]
STYQTMPPR [▼] GMEPHMSEC		+		Ber e 1	[28]
EEGVSLEKR [▼] <i>ea</i> _Δ EAPYGRG	+		+	SFA8	[28]
GERSQVSPR [▼] MREEDHKQL	+			SFA8	[28]

All the sites are deduced from the results of N-terminal sequencing and mass spectrometry. Kex2 or Yapsin family endopeptidase activity is shown as (*), while symbols (_Δ) and (_□) indicate where the activities of dipeptidyl aminopeptidase (Ste13) and carboxypeptidase (Kex1), respectively, actually stop. Small print italics=residues possibly removed by exopeptidases.

(a) rAL1



(b) rAL3



(c) nAL3

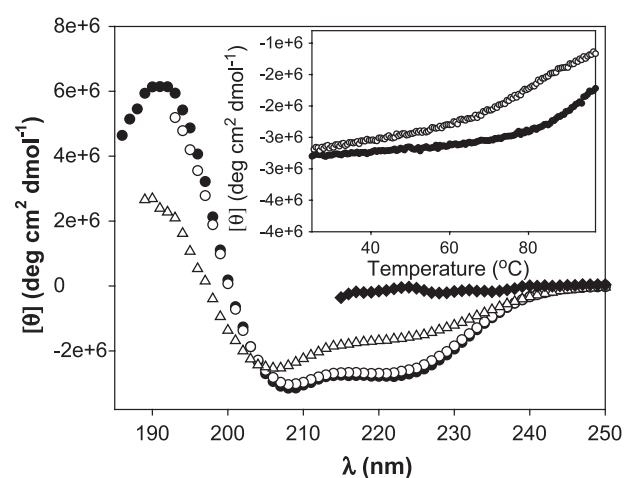


Fig. 4. CD spectra of recombinant and native soybean 2S albumins under various conditions in 10 mM phosphate buffer pH 6.8. (●) 25 °C; (Δ) 97 °C; (○) 10 mM glycine buffer pH 2.2; and (◆) 6 M guanidinium chloride. Insets: heat-induced unfolding transition of proteins in 10 mM phosphate buffer pH 6.8 (●) and 10 mM glycine buffer pH 2.2 (○), followed by far UV CD measurements at 222 nm.

3.5. Secondary structure of the soybean 2S albumins

The secondary structures of both nAL and rALs were predicted by the consensus secondary structure prediction (Jpred) [29] and the profile network prediction (SUBsec) [30,31] (Fig. 2). According to the predicted secondary structure, the soybean 2S albumins contained a large proportion of α -helical structures.

Far UV CD spectroscopy of the native and recombinant albumins under non-denaturing conditions (10 mM phosphate buffer, pH 6.8, 25 °C) showed that they had similar structures and confirmed the presence of high contents of α -helix (Fig. 4). Similar spectra of these three albumins were obtained at 10 mM glycine buffer pH 2.2, 25 °C, at phosphate buffer pH 6.8, 97 °C and in 6 M guanidinium chloride, with the latter two conditions resulting in the partial loss of α -helical structure (Fig. 4). However, further studies demonstrated that rAL1 exhibited similar stability to thermal denaturation at pH 2.2 and at pH 6.8, whereas nAL3 and rAL3 were more stable to thermal denaturation at pH 6.8 (Fig. 4, inset). This indicates that nAL3 and rAL3 have essentially identical structures despite their different processing sites.

4. Discussion

2S albumins from a wide range of dicotyledonous species constitute a family of structurally related proteins, which typically consist of two subunits joined by disulfide bonds with highly conserved positions of cysteine residues [7]. Although the major role in vivo of 2S albumins is thought to be as seed storage proteins, several other types of biological activity have been reported, including allergenicity to human [8,32], trypsin-inhibition [33], calmodulin antagonism [34], and antifungal activity [35].

The first 2S albumin to be isolated from soybean was a 4.4 kDa aspartic acid-rich peptide, which actually corresponds to the small subunit of AL3 [36]. The presence of a poly-aspartic acid sequence and a cell attachment sequence (Arg–Gly–Asp) raised an interesting question as to whether it had any role in seed development. Further interest in its biological function was raised by the antimetabolic effect of this peptide when transiently expressed in mammalian cells [12]. The cDNAs encoding the sequences of the aspartic acid-rich 2S albumin (AL3) and another 2S albumin (AL1), which are 74.1% identical in sequence, have subsequently been submitted to GenBank.

Initial screening results using soybean-sensitive patients have not so far implicated 2S albumins as major allergens [15,16,37], but a recent study of soy lecithin has suggested that the large subunit of AL3 might be recognized by patients' serum IgE [38].

In order to study their structure–function relationship, we have expressed two soybean 2S albumins in the methylophilic yeast *P. pastoris* expression system. The high

yields and the ability to carry out post-translational modifications of eukaryotic proteins have made the *P. pastoris* expression system very appealing [39].

Using the α -factor secretion signal and growing the cells in minimal medium, the rALs were secreted at around 4.5 mg/l at 80 h of induction, representing more than 20% of the total secreted proteins. By analogy with other systems [39], use of a more complex medium combined with cultivation at high cell densities, would be expected to provide even higher protein yields.

Many proteins expressed in *P. pastoris* are glycosylated [39] and both rAL1 and rAL3 have consensus N-glycosylation sites. Structure prediction analysis indicated that no exposed Ser and Thr, potential sites of *O*-glycosylation, are present in loop regions. The only post-translational modification detected during the secretion of rALs from *P. pastoris* was proteolytic processing. All of the masses obtained from mass spectra could be deduced from the amino acid sequences, indicating that no glycans were attached to either the native or the recombinant albumins.

In common with other 2S albumins [11], the precursor nAL3 expressed in the plant seed is proteolytically processed into a 5.1 kDa small subunit and a 9.0 kDa large subunit with the removal of the signal peptide and a linker peptide between the two subunits. We used the *S. cerevisiae* α -factor signal sequence to direct the rAL protein into the secretory pathway of *P. pastoris*. Endoproteolytic removal of the signal sequence by the *P. pastoris* Kex2 protease and trimming of the cleavage products by Ste13 aminopeptidase would therefore be expected to give the mature rAL with the correct N-terminus.

Plant cells, different from *P. pastoris*, may have a machinery for delivery of protein aggregates produced within the ER to the vacuoles bypassing the Golgi apparatus [40]. In plants, the proteolytic processing of storage proteins (11S, 7S, and 2S) occurs in the vacuole catalysed by a specific class of cysteinyl proteases called legumains [41]. Although it is unlikely that they are related to the processing enzymes in *P. pastoris*, the specificity for the same loop region of the 2S albumins is clearly similar.

In *P. pastoris* the secretion of two (rAL1) or three (rAL3) small subunits of rAL varying in their C-termini was observed, indicating cleavages at several internal processing sites within the same loop region. A possible explanation is that initial endoproteolytic cleavages adjacent to monobasic or dibasic amino acids was followed by less efficient processing by carboxypeptidases.

Several endoproteases located in the secretory pathway of yeast have been described and are known to be involved in protein processing. The Kex2 enzyme is generally considered to be responsible for the initial processing of both yeast and foreign proteins at dibasic sites [42,43], but in an increasing number of cases, this enzyme has also been found to be involved in cleavage at single Arg residues [44,45]. Other possible candidate endoproteases are members of a

recently discovered family of aspartic proteases, the yapsins, which are capable of cleaving at dibasic and monobasic residues, e.g. single Arg residue. Three members of this family, with high sequence similarity, were identified in *S. cerevisiae*: yapsin 1 (also known as Yap3p) [46,47], yapsin 2 (also known as Mkc7p) [48] and yapsin 3 [49]. Although their specificities are slightly different, these enzymes are all able to process α -factor pheromone when this pheromone is over-expressed in Kex2-deficient mutants. Therefore, both the Kex2 protease and members of the yapsin family may contribute to the internal processing.

The Kex1 carboxypeptidase, which is located in the Golgi apparatus and involved in the removal of the C-terminal LysArg dipeptide of α -factor pheromone, may be involved in the C-terminal degradation of the small subunits of rAL. A previous report on the expression of recombinant desulfato-hirudin expressed in *S. cerevisiae* showed that this protease was directly involved in the trimming of C-terminal LeuGln residues, indicating that it has a broader substrate spectrum than basic residues [50]. Furthermore, the removal of a large number of residues from the rAL suggests that lysis of yeast cells occurred during fermentation for 80 h, releasing vacuolar proteases with broad substrate specificity, such as carboxypeptidase Y, into the culture supernatant.

One form of the large subunit of rAL1 had the N-terminus RKKE, demonstrating cleavage between Ile⁵⁶Arg⁵⁷ (rAL1). We do not know if this resulted from inaccurate cleavage by the endoproteases mentioned above, or cleavage by a currently uncharacterized protease. Since inaccurate cleavage by the Kex2 protease or the yapsin family proteases has not previously been reported, we favour the latter possibility.

Interestingly, although some other monobasic and dibasic residues are present in the sequence of rAL, such as Arg³³Arg³⁴ (rAL1) and Ile¹¹⁸Arg¹¹⁹ (rAL1), they were not cleaved. This suggests that protein conformation is an additional factor determining the specificity of recognition by the relevant proteases. According to the predicted secondary structures, the cleavage sites in the native and recombinant proteins were located in extended turn or loop structures; these regions would be flexible and hence more accessible for proteolytic processing. Conversely, unprocessed sequences are predicted to be present in α -helical structures, which may be poor substrates for proteolytic enzymes, especially when the proteases are membrane-bound (e.g. Kex2) and may have additional steric constraints.

The potential intrinsic allergenicity of proteins might be related to their higher physicochemical stability [51]. Therefore, CD spectroscopy was used to assess the stability of both soybean 2S albumins to denaturation by temperature and guanidinium chloride. Consistent with the three-dimensional structures determined for 2S albumins from other species [11], they generally contained a large proportion of α -helical secondary structures. Both recombinant and native proteins retained most of their secondary structure at temperatures as high as 97 °C. All three

proteins exhibited high resistance to chemically induced unfolding, retaining a significant amount of secondary structure in concentrations of guanidinium chloride as high as 6 M. The differences between the internal cleavage sites of rAL3 and nAL3 could relate to small local differences in conformation, although CD spectroscopy showed little difference between the structure and stability of the whole proteins.

In conclusion, the 2S albumins from soybean are unusually rich in charged residues and possess similar properties to the widespread 2S albumin family of proteins. They possess the distinctive disulfide structure of the class, are non-glycosylated, and are very stable to temperature and chemical denaturants. The folded proteins (AL1 and AL3) are well expressed by the methylotrophic yeast *P. pastoris* and the subunits are post-translationally processed within the same loop region as the native protein but with some differences in the precise sites. Despite the wide consumption of soybean seeds there are no reported cases of food allergy to the 2S albumin fraction. The recombinant proteins reported in this study will therefore allow this aspect to be studied in more detail as well as providing material for studies of structure–function relationships.

Acknowledgements

J. Lin is supported by a Full Tuition Fee Scholarship from the University of Nottingham. We would also like to thank Dr. Rudolph Jung and Pioneer Hi-Bred International, Inc. for making available the method for purification of the native soybean 2S albumin, the plasmids containing the cDNA encoding soybean 2S albumins and one of the antibodies used in this work.

References

- [1] T. Ogawa, E. Tayama, K. Kitamura, N. Kaizuma, Genetic improvement of seed storage proteins using 3 variant alleles of 7S globulin subunits in soybean (*Glycine max* L.), Jpn. J. Breed. 39 (1989) 137–147.
- [2] J.F. Thompson, J.T. Madison, M.A. Waterman, A.M.E. Muenster, Effect of methionine on growth and protein-composition of cultured soybean cotyledons, Phytochemistry 20 (1981) 941–945.
- [3] S.K. Dhir, S. Dhir, M.A. Savka, F. Belanger, A.L. Kriz, S.K. Farrand, J.M. Widholm, Regeneration of transgenic soybean (*Glycine max*) plants from electroporated protoplasts, Plant Physiol. 99 (1992) 81–88.
- [4] D.E. McCabe, W.F. Swain, B.J. Martinell, P. Christou, Stable transformation of soybean (*Glycine max*) by particle-acceleration, Bio-Technol. 6 (1988) 923–926.
- [5] M.J. Revilleza, A.F. Galvez, D.C. Krenz, B.O. deLumen, An 8 kDa methionine-rich protein from soybean (*Glycine max*) cotyledon: identification, purification, and N-terminal sequence, J. Agric. Food Chem. 44 (1996) 2930–2935.
- [6] R.J. Youle, A.H.C. Huang, Occurrence of low-molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species, Am. J. Bot. 68 (1981) 44–48.

- [7] P.R. Shewry, M.J. Pandya, The 2S albumin storage proteins, in: P.R. Shewry, R. Casey (Eds.), *Seed Proteins*, Kluwer Academic, Netherlands, 1999, pp. 619–643.
- [8] J.A. Nordlee, S.L. Taylor, J.A. Townsend, L.A. Thomas, R.K. Bush, Identification of a brazil nut allergen in transgenic soybeans, *N. Engl. J. Med.* 334 (1996) 688–692.
- [9] L. Lo Conte, S.E. Brenner, T.J.P. Hubbard, C. Chothia, A.G. Murzin, SCOP database in 2002: refinements accommodate structural genomics, *Nucleic Acids Res.* 30 (2002) 264–267.
- [10] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP—a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* 247 (1995) 536–540.
- [11] M. Rico, M. Bruix, C. Gonzalez, R.I. Monsalve, R. Rodriguez, ¹H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein, *Biochemistry* 35 (1996) 15672–15682.
- [12] A.F. Galvez, B.O. de Lumen, A soybean cDNA encoding a chromatin-binding peptide inhibits mitosis of mammalian cells, *Nat. Biotechnol.* 17 (1999) 495–500.
- [13] M. Hellerstein, D.H. Calloway, Antimitotic peptide characterized from soybean: role in protection from cancer? *Nutr. Rev.* 57 (1999) 359–361.
- [14] A.R. Kennedy, The evidence for soybean products as cancer preventive agents, *J. Nutr.* 125 (1995) S733–S743.
- [15] R. Gonzalez, L. Zapatero, F. Caravaca, J. Carreira, Identification of soybean proteins responsible for respiratory allergies, *Int. Arch. Allergy Appl. Immunol.* 95 (1991) 53–57.
- [16] A.M. Herian, S.L. Taylor, R.K. Bush, Identification of soybean allergens by immunoblotting with sera from soy-allergic adults, *Int. Arch. Allergy Appl. Immunol.* 92 (1990) 193–198.
- [17] L.A. Moroz, W.H. Yang, Kunitz soybean trypsin inhibitor: a specific allergen in food anaphylaxis, *N. Engl. J. Med.* 301 (1980) 1126–1128.
- [18] R.K. Bush, S.L. Taylor, J.A. Nordlee, W.W. Busse, Soybean oil is not allergenic to soybean-sensitive individuals, *J. Allergy Clin. Immunol.* 76 (1985) 242–245.
- [19] T. Ogawa, N. Bando, H. Tsuji, K. Nishikawa, K. Kitamura, Alpha-subunit of beta-conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic-dermatitis, *Biosci. Biotechnol. Biochem.* 59 (1995) 831–833.
- [20] T. Ogawa, H. Tsuji, N. Bando, K. Kitamura, Y.L. Zhu, H. Hirano, K. Nishikawa, Identification of the soybean allergenic protein, Gly m Bd 30 k, with the soybean seed 34-kDa oil-body-associated protein, *Biosci. Biotechnol. Biochem.* 57 (1993) 1030–1033.
- [21] H. Tsuji, N. Bando, M. Hiemori, R. Yamanishi, M. Kimoto, K. Nishikawa, T. Ogawa, Purification and characterization of soybean allergen Gly m Bd 28 K, *Biosci. Biotechnol. Biochem.* 61 (1997) 942–947.
- [22] M.G. Zeece, T.A. Beardslee, J.P. Markwell, G. Sarath, Identification of an IgE-binding region in soybean acidic glycinin G1, *Food Agric. Immunol.* 11 (1999) 83–90.
- [23] H.P. Rihs, Z.P. Chen, F. Rueff, A. Petersen, P. Rozynek, H. Heimann, X. Baur, IgE binding of the recombinant allergen soybean profilin (rGly m 3) is mediated by conformational epitopes, *J. Allergy Clin. Immunol.* 104 (1999) 1293–1301.
- [24] J. Kleins-Tebbe, A. Wangorsch, L. Vogel, D.N. Crowell, U.F. Hausstein, S. Vieths, Severe oral allergy syndrome and anaphylactic reactions caused by a Bet v 1-related PR-10 protein in soybean, SAM22, *J. Allergy Clin. Immunol.* 110 (2002) 797–804.
- [25] S. Odani, T. Koide, T. Ono, Y. Seto, T. Tanaka, Soybean hydrophobic protein-isolation, partial characterization and the complete primary structure, *Eur. J. Biochem.* 162 (1987) 485–491.
- [26] R. Codina, R.F. Locket, E. FernandezCaldas, R. Rama, Purification and characterization of a soybean hull allergen responsible for the Barcelona asthma outbreaks: 2. Purification and sequencing of the Gly m 2 allergen, *Clin. Exp. Allergy* 27 (1997) 424–430.
- [27] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.
- [28] M.J.C. Alcocer, G.J. Murtagh, K. Bailey, M. Dumoulin, A. Sarabia, M.J. Parker, D.B. Archer, The disulphide mapping, folding and characterisation of recombinant *Ber e 1* and SFA-8, two sulphur-rich 2S plant albumin allergens, *J. Mol. Biol.* 324 (2002) 165–175.
- [29] J.A. Cuff, M.E. Clamp, A.S. Siddiqui, M. Finlay, G.J. Barton, Jpred: a consensus secondary structure prediction server, *Bioinformatics* 14 (1998) 892–893.
- [30] B. Rost, C. Sander, Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol.* 232 (1993) 584–599.
- [31] B. Rost, C. Sander, Combining evolutionary information and neural networks to predict protein secondary structure, *Proteins* 19 (1994) 55–72.
- [32] J.D. Kelly, S.L. Hefle, 2S methionine-rich protein (SSA) from sunflower seed is an IgE-binding protein, *Allergy* 55 (2000) 556–559.
- [33] N. Genov, I. Goshev, D. Nikolova, D.N. Georgieva, B. Filippi, I. Svendsen, A novel thermostable inhibitor of trypsin and subtilisin from the seeds of *Brassica nigra*: amino acid sequence, inhibitory and spectroscopic properties and thermostability, *Biochim. Biophys. Acta, Prot. Struct. Mol. Enzymol.* 1341 (1997) 157–164.
- [34] G.M. Polya, S. Chandra, R. Condrón, Purification and sequencing of radish seed calmodulin antagonists phosphorylated by calcium-dependent protein-kinase, *Plant Physiol.* 101 (1993) 545–551.
- [35] F.R.G. Terras, H.M.E. Schoofs, M.F.C. Debolle, F. Vanleuven, S.B. Rees, J. Vanderleyden, B.P.A. Cammue, W.F. Broekaert, Analysis of 2 novel classes of plant antifungal proteins from radish (*Raphanus sativus* L) seeds, *J. Biol. Chem.* 267 (1992) 15301–15309.
- [36] S. Odani, T. Koide, T. Ono, Amino acid sequence of a soybean (*Glycine max*) seed polypeptide having a poly(L-aspartic acid) structure, *J. Biol. Chem.* 262 (1987) 10502–10505.
- [37] T. Ogawa, N. Bando, H. Tsuji, H. Okajima, K. Nishikawa, K. Sasaoaka, Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis, *J. Nutr. Sci. Vitaminol.* 37 (1991) 555–565.
- [38] X.L. Gu, T. Beardslee, M. Zeece, G. Sarath, J. Markwell, Identification of IgE-binding proteins in soy lecithin, *Int. Arch. Allergy Immunol.* 126 (2001) 218–225.
- [39] J.L. Cereghino, J.M. Cregg, Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *FEMS Microbiol. Rev.* 24 (2000) 45–66.
- [40] I. Hara-Nishimura, T. Shimada, K. Hatano, Y. Takeuchi, M. Nishimura, Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles, *Plant Cell* 10 (1998) 825–836.
- [41] K. Muntz, A.D. Shutov, Legumains and their functions in plants, *Trends Plant Sci.* 7 (2002) 340–344.
- [42] C. Brenner, R.S. Fuller, Structural and enzymatic characterization of a purified prohormone-processing enzyme-secreted, soluble Kex2 protease, *Proc. Natl. Acad. Sci. U. S. A.* 89 (1992) 922–926.
- [43] E.C. Ledgerwood, P.M. George, R.J. Peach, S.O. Brennan, Endoproteolytic processing of recombinant proalbumin variants by the yeast kex2 protease, *Biochem. J.* 308 (1995) 321–325.
- [44] I.J. Rourke, A.H. Johnsen, N.N. Din, J.G.L. Petersen, J.F. Rehfeld, Heterologous expression of human cholecystokinin in *Saccharomyces cerevisiae*—evidence for a lysine-specific endopeptidase in the yeast secretory pathway, *J. Biol. Chem.* 272 (1997) 9720–9727.
- [45] Y.S. Zhu, X.Y. Zhang, C.P. Cartwright, D.J. Tipper, Kex2-dependent processing of yeast K(1) killer preprotoxin includes cleavage at proarg-44, *Mol. Microbiol.* 6 (1992) 511–520.
- [46] A.V. Azaryan, M. Wong, T.C. Friedman, N.X. Cawley, F.E. Estivariz, H.C. Chen, Y.P. Loh, Purification and characterization of a paired basic residue-specific yeast aspartic protease encoded by the yap3 gene—similarity to the mammalian pro-opiomelanocortin-converting enzyme, *J. Biol. Chem.* 268 (1993) 11968–11975.
- [47] N.X. Cawley, H.C. Chen, M.C. Beinfeld, Y.P. Loh, Specificity and kinetic studies on the cleavage of various prohormone mono- and

- paired-basic residue sites by yeast aspartic protease 3, *J. Biol. Chem.* 271 (1996) 4168–4176.
- [48] H. Komano, R.S. Fuller, Shared functions in vivo of a glycosyl-phosphatidylinositol-linked aspartyl protease, Mkc7, and the proprotein processing protease Kex2 in yeast, *Proc. Natl. Acad. Sci. U. S. A.* 92 (1995) 10752–10756.
- [49] V. Olsen, N.X. Cawley, J. Brandt, M. Egel-Mitani, Y.P. Loh, Identification and characterization of *Saccharomyces cerevisiae* yapsin 3, a new member of the yapsin family of aspartic proteases encoded by the YPS3 gene, *Biochem. J.* 339 (1999) 407–411.
- [50] J. Heim, K. Takabayashi, B. Meyhack, W. Marki, G. Pohlig, C-terminal proteolytic degradation of recombinant desulfato-hirudin and its mutants in the yeast *Saccharomyces cerevisiae*, *Eur. J. Biochem.* 226 (1994) 341–353.
- [51] J.D. Astwood, J.N. Leach, R.L. Fuchs, Stability of food allergens to digestion in vitro, *Nat. Biotechnol.* 14 (1996) 1269–1273.
- [52] B. Rost, C. Sander, Conservation and prediction of solvent accessibility in protein families, *Proteins* 20 (1994) 216–226.