



Expression in *Escherichia coli* and disulfide bridge mapping of PSC33, an allergenic 2S albumin from peanut

Gilles Clement ^{a,*}, Didier Boquet ^b, Lucie Mondoulet ^a, Patricia Lamourette ^b,
Hervé Bernard ^a, Jean Michel Wal ^a

^a Laboratoire INRA-CEA d'immunoallergie alimentaire, SPI Bât 136 CEA, Saclay 91191, Gif sur Yvette Cedex, France

^b Service de Pharmacologie et d'Immunologie, SPI Bât 136 CEA, Saclay 91191, Gif sur Yvette Cedex, France

Received 10 March 2005, and in revised form 24 May 2005

Abstract

In this work, we describe the expression, purification, and disulfide mapping of the named 'peanut seed cDNA 33' (PSC33) peanut allergen. A variant of PSC33 (with N₆₃, E₆₄, Q₆₉ instead of D₆₃, Q₆₄, E₆₉) has been identified in peanut by proteomic analysis of a highly IgE immunoreactive purification fraction. It is 92% homologous to Ara h 6. We raised monoclonal antibodies against PSC33 and amplified it by PCR from peanut leaf genomic DNA. PSC33 was intron-less and the two NEQ and DQE variants of PSC33 were equally amplified. Since expression of the natural PSC33 (DQE) gene was very low in *Escherichia coli* even with supplementation of rare codon tRNAs, a synthetic gene optimized for expression in *E. coli* of PSC33 (DQE) was introduced into a pET9-c vector. A high production of protein occurred in the inclusion bodies that was submitted to refolding using an additive-introduced stepwise dialysis protocol which consists in the gradual removal of the denaturing agent guanidine-HCl with controlled introduction of oxidized and reduced glutathione and L-arginine as a chemical chaperone. After reverse phase HPLC purification, 1 mg of pure refolded protein (as assayed by MALDI-TOF mass spectrometry, mouse IgG immunoreactivity and circular dichroism) were obtained with every 100 ml of bacterial culture. Trypsin and CNBr hydrolysis of the protein combined with MALDI-TOF mass spectrometry allowed us to assign disulfide bridges and show that the native and refolded proteins were identical. The four disulfides of canonical 2S albumins were conserved and the two supplementary cysteines of PSC33 were paired together.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Peanut allergen; Ara h 6; 2S albumin; *Escherichia coli*; Refolding; Additive-introduced stepwise dialysis; Disulfide bridge mapping

2S albumins are seed storage proteins found in dicotyledonous plants and particularly in legumes. They share with cereals α -amylase/trypsin inhibitors and non-specific lipid transfer proteins (nsLTP)¹ a characteristic four helix, four disulfide bridges structure which has proved of great utility in plant evolution by providing respec-

tively food for the embryo, protection from predators, and lipid transfer [1,2]. This structure provides compactness and resistance to proteolysis and this might be why a high number of plant allergens have it. The prolamin superfamily to which these three subfamilies belong counts 39 members out of 133 plant allergens recently compiled and as such is the bigger family of plant allergens [3].

Nine peanut allergens have been cloned (Ara h 1–8 and oleosin) and the corresponding proteins have all been isolated except Ara h 7 [4]. Not only the 2S is the albumin family represented (Ara h 2, 6, 7), but also 7S globulins (Ara h 1), 11S globulins (Ara h 3, 4), profilin

* Corresponding author. Fax: +33 1 69 08 59 07.

E-mail address: gilles.clement@cea.fr (G. Clement).

¹ Abbreviations used: nsLTP, non-specific lipid transfer proteins; CHCA, α -cyano-hydroxy-cinnamic acid; TFA, trifluoroacetic acid; IPTG, isopropylthio- β -D-galactoside; DTT, dithiothreitol; Gu-HCl, guanidine-HCl; CD, circular dichroism.

(Ara h 5), and Bet v 1 (Ara h 8). PSC33 (AF366561) has not been described as an allergen but as a gene differentially expressed in peanut seed development: it is not expressed before 40 days after pollination, together with an other 2S albumin PSC32 which is the long variant of Ara h 2 [5]. PSC33 differs from Ara h 6 (AF092846) by only 10 amino acids (92% homology). Twenty-five 2S albumins allergens were extracted from the 2206 entries of the allergome database (www.allergome.org) among which Sin a I and Bra J Ie (mustard), Ber e 1 (brazil nut), Ric c 1 and 3 (castor bean), Ses i 1 and 2 (sesame), Jug r 1 (English walnut), and Cic a (chickpea).

Three solution structures of 2S albumins have been determined by NMR spectroscopy: rproBnIb (napin), Hel a 2S albumin (SFA-8), and Ric c 3 [6–8], and the structural determination of Ara h 6 [9] is in progress. These structures are invaluable tools for those like us who seek to elucidate allergenic epitopes [10]. The napin structure was recently used to model Ber e 1 whose potential structural epitopes were then substituted on the analogous SFA-8 used as a platform [11]. These mutagenesis approaches are very promising but require large amounts of purified and properly refolded proteins. Several methodologies are used to achieve this goal. Expression in *Pichia pastoris* yields refolded proteins secreted in the culture medium (Ber e 1, SFA-8, and rproBnIb) [11,7,6]. Expression as a fusion protein in *Escherichia coli* (Sin a 1, Ara h 2, and Ara h 6) [12,13]. Ric c 3 has also been expressed in *E. coli* grown in a defined culture medium and is secreted in the medium with proper refolding [14]. Finally the protein accumulated in *E. coli* inclusion bodies can be solubilized and refolded in vitro (wheat α -amylase inhibitor CM16 and 0.19) [15,16]. We chose this last method because we have experienced in our laboratory the refolding of two different single-chain antibody fragment (scFv) using a stepwise dialysis method [17] and like the authors of [16] we decided to make a synthetic gene of PSC33 (DQE) to circumvent its high number of codons of low usage in *E. coli* [18].

Although cloning the natural gene was of no use for PSC33 expression, it has confirmed the existence of two variants of PSC33 already deposited in the NCBI protein database. The good yield of the synthetic gene expression allowed us to map the disulfide bridges of PSC33. This work shows that expression in *E. coli* can still be useful for obtaining high amounts of properly folded proteins.

Materials and methods

Proteomic analysis

The proteins from commercial roasted peanuts bought in France (Virginia variety) were fractionated by ammonium sulfate precipitation, ion-exchange, and reverse phase chromatography (H. Bernard, manuscript in preparation). IgE immunoreactive fractions were sub-

mitted to one-dimensional SDS–PAGE. Proteins were in-gel digested and carbamidomethylated according to the protocol found at www.narrador.embl-heidelberg.de/ on the EMBL bioanalytical research group site (in the directory ‘activities’). Digestion with porcine trypsin (Promega) at 40 ng/ μ l was performed for 30 min at 58 °C [19]. α -Cyano-hydroxy-cinnamic acid (CHCA) (Sigma-Aldrich) was recrystallized in boiling ethanol and used saturated in 50% CH₃CN in water containing 0.3% trifluoroacetic acid (TFA). MALDI-TOF spectra were acquired with a Voyager DE-RP instrument (Applied Biosystems) for the identification of PSC33 in peanut extract and with a Voyager DE-STR (Applied Biosystems) for the disulfide mapping. Databases were searched online with the search engines Profound and Protein Prospector available at www.expasy.ch.

Genomic cloning

Peanut seeds from Reunion island (Indian Ocean) were grown in the laboratory. Half a leaf (about 60 mg) was ground dry in a Ribolyzer for 2 $\times$ 15 s and then resuspended in 400 μ l of AP1 buffer, the first step of the DNeasy Plant mini DNA extraction kit (Qiagen). Five micrograms of peanut genomic DNA were obtained. PCR amplification was done on 50–250 ng of genomic DNA with AccuTaq DNA polymerase (Sigma) for the following 25 cycles: 30 s denaturation at 94 °C, 30 s hybridization at 65–68 °C depending on the primer, 60 s extension at 68 °C. The first cycle was preceded by 60 s denaturation at 94 °C and the last cycle followed by 7 min extension at 68 °C. For amplification in the propeptide, the following forward primer was used: 5'GCC AAG TCC ACC ATC CTG GTA GCC C3'. For amplification starting at the N-terminal of the mature protein, the following forward primer was used: 5'GCG ATG AGG CGC GAG AGG GGG CGA CAA GGG G3'. The following reverse primer was used in both cases: 5'GCA TCT GCC GCC ACT CAC GTC CAA ATC GCA ACG CTG TGG TGC3'. The PCR products were cloned directly in pTrcHis2-TOPO, pTrcHis-TOPO, and pCR4-TOPO vectors (Invitrogen). Plasmids were purified with the Bio-Rad plasmid mini kit and sequenced in both directions by MWG-Biotech (Ebersberg, Germany). The natural gene in pTrcHis2-TOPO and pTrcHis-TOPO was screened for expression in TOP10 (Invitrogen) and BL21-Codonplus-RIL (Stratagene) *E. coli*. It was also resubmitted to PCR for cloning into pET100D-TOPO and expression screening in BL21 (DE3) and BL21-Codon plus (DE3)-RIL (Stratagene).

Synthesis of a PSC33 gene optimized for expression in *E. coli*

PSC33 gene was assembled from nine forward primers and nine reverse primers ranging from 33 to 48

nucleotides and synthesized by MWG-Biotech (Germany) (Fig. 3). They were all 5'-phosphorylated with the exception of the first forward and the last reverse. Twelve picomoles of each primer were mixed in 50 μ l of 1 $\times$ NEB buffer 1 (New England Biolabs). The primer mixture was warmed to 90 °C for 5 min in a 4 L water bath. Heating was stopped and the temperature decreased slowly to 39 °C in 3 h. The tube was then maintained at 37 °C for 10 min before the addition of 10 $\times$ T4 DNA ligase buffer and 2 μ l of undiluted T4 DNA ligase (New England Biolabs). Ligation lasted 30 min after which the ligase was inactivated by heating for 10 min at 65 °C. The gene was then amplified by PCR with AccuTaq DNA polymerase using 20-mer forward and reverse primers and the previously described PCR conditions. After standing overnight at 4 °C, the PCR product was then cloned in pTrcHis2-TOPO and pCR4-TOPO vectors. Sixty clones in pTrcHis2 and 36 clones in pCR4 were checked for the presence of the insert by direct PCR on the clones grown for 3.5 h in 200 μ l of LB medium and 24 out of 60 and 12 out of 36 were positive respectively on a 96-track agarose gel (E-gel, Invitrogen). Nine clones were sent for sequencing.

Cloning of the synthetic gene into pET100D-TOPO, pET3c, and pET9c vectors

The synthetic gene was introduced in the pET100D-TOPO vector by PCR cloning according to the manufacturer's instructions. For cloning into pET3c and pET9c vectors (Novagen), *Nde*I and *Bam*HI sites were first added 5' and 3' of the gene, respectively, by PCR and cloning into pTrcHis-TOPO vector. Sixteen micrograms of a purified plasmid containing the insert and separately 10 μ g of pET3c or pET9c were then double digested overnight at 37 °C with 50 U of each enzyme in 50 μ l of 1 $\times$ NEB buffer *Bam*HI. The vector was then dephosphorylated with 10 U of calf intestinal phosphatase (NEB) for 30 min at 37 °C while the insert was left phosphorylated. Fragments were separated by electrophoresis on a 1% agarose gel and bands were purified from the gel using the Spin prep Gel DNA kit (Novagen) but replacing the binding columns from this kit with those of the Bio-Rad plasmid mini kit. Eluted DNA was quantified spectrophotometrically using UVettes (Eppendorf). Five microlitres of vector (235 ng) were mixed with 5 μ l of insert (45 ng), 1 μ l of ligase buffer, and 1 μ l of T4 DNA ligase, and incubated for 30 min at 20 °C. The ligated plasmid was then transformed in chemocompetent BL21 (DE3) cells (Novagen) for expression screening.

Expression of PSC33

Escherichia coli cells were grown overnight at 37 °C in Miller's Luria broth (LB) (Sigma). They were then seeded at 2% vol/vol in fresh medium and grown until the optical density (OD) at 600 nm reached 0.6 (in 105 min). An ali-

quot was removed for the non-induced control and the rest of the culture was supplemented with isopropylthio- β -D-galactoside (IPTG) (Invitrogen) at concentrations 0.4 mM for pET3c and pET9c vectors or 1 mM for the other vectors. Culture volumes varied from 10 ml for screening to 500 ml for production. Induction was allowed to proceed for 5 h. One millilitre aliquots of non-induced and induced cells were collected for electrophoretic verification of the expression. Bacteria were centrifuged and the pellet was stored at -20 °C until further processing. The 1 ml aliquots were resuspended in 100 μ l of 7 M urea, 100 mM Tris, pH 8, 10 mM dithiothreitol (DTT), heated for 5 min at 95 °C and centrifuged. Supernatants were analysed with PhastSystem (Amersham Biosciences): 1 vol of 5 $\times$ concentrated sample buffer was added to 4 vol of supernatant and reheated for 5 min at 95 °C. One microlitre of these samples was analysed on 10–15% acrylamide gradient gels and Coomassie blue stained.

PSC33 refolding and purification

Frozen cell pellets were resuspended in 10 ml of 0.1 M Tris, pH 8, containing a cocktail of protease inhibitors (complete mini, Roche) for 100 ml of bacterial culture. They were sonicated for 45 s at room temperature and centrifuged for 10 min at 12,000g. The inclusion bodies' pellet was then resuspended with the same volume ratio in 0.1 M Tris, pH 8, 6 M guanidine-HCl (Gu-HCl), 5 mM DTT, and rotated for 1–2 h on a rotary mixer. The extract was centrifuged at 12,000g for 10 min, and the supernatant was put in a dialysis cassette of MWCO 3500 (Pierce) and submitted to the refolding protocol of Tsumoto et al. [17]. The concentration of Gu-HCl in the dialysis bath (1 L of 0.1 M Tris, pH 8, for a 10 ml cassette) was decreased in the following steps: 6 M $\rightarrow$ 3 M $\rightarrow$ 2 M $\rightarrow$ 1 M $\rightarrow$ 0.5 M $\rightarrow$ 0 M. Step changes were applied overnight for 6 M, over daytime for 3 M, overnight for 2 M, and so on. Oxidized glutathione (250 μ M) and reduced glutathione (500 μ M) were added at the 1 M Gu-HCl step and readded with 0.4 M arginine at the 0.5 M Gu-HCl step. Precipitation occurred in the last step but was not due to PSC33. The refolded protein was then applied to a semipreparative (10 $\times$ 250 mm) C18 (10 μ m, 300 Å) reverse phase HPLC column (Vydac) operated at 3 ml/min and 40 °C with a Waters 600 pump. Buffer A was 0.1% TFA in water, Buffer B was 100% CH₃CN + 0.04% TFA. One-minute fractions were analysed by MALDI-TOF, freeze-dried individually, weighed and analysed by SDS-PAGE, circular dichroism and immunoassay.

Enzyme immunoassay

Monoclonal antibodies directed against native PSC33 and native PSC33 coupled to acetylcholinesterase (PSC33-AChE) were obtained according to the procedures described by Negroni et al. [20]. Competitive immunoassay procedures are described by Clement et al. [10].

Briefly, 50 μ l of raw hybridoma culture supernatant, 50 μ l of PSC33–AChE (1 Ellman unit per millilitre), and 50 μ l of purified protein fractions at various concentration were co-incubated overnight at 4°C in microtitration plates coated with 5 μ g/ml of a goat anti-mouse polyclonal antibody (Jackson, West Grove PA). Plates were washed five times with 0.01 M phosphate buffer, pH 7.4, containing 0.05% Tween 20, filled with 200 μ l of Ellman's reagent and read at 414 nm after 10 min.

Circular dichroism measurement

Circular dichroism (CD) spectra were recorded at room temperature with a Jobin Yvon CD3 dichrograph in a 0.05 cm cell (200 μ l were needed to completely fill it). Purified PSC33 was 100 μ g/ml in 10 mM phosphate buffer, pH 7.4. Non-refolded purified

recombinant PSC33 was obtained from direct injection of an induced *E. coli* Gu–HCl extract on a Vydac C18 analytical HPLC column, speed-vac drying of the fractions and resuspension at 100 μ g/ml in 10 mM phosphate buffer, pH 7.4. UV absorptions were checked before CD analysis. Spectra were recorded from 190 to 260 nm with 0.5 nm steps. Integration time was 1 s with a slit width of 0.1 nm and a 2 nm constant bandpass. Three spectra were averaged. The α -helix contents were estimated with the program K2d (<http://www.embl-heidelberg.de/~andrade/k2d/>) and molar ellipticity at 222 nm [28,29].

Disulfide bridge assignment

Models were drawn using the GPMW 5.11 software. Refolded PSC33 was digested in aliquots of 1 mg

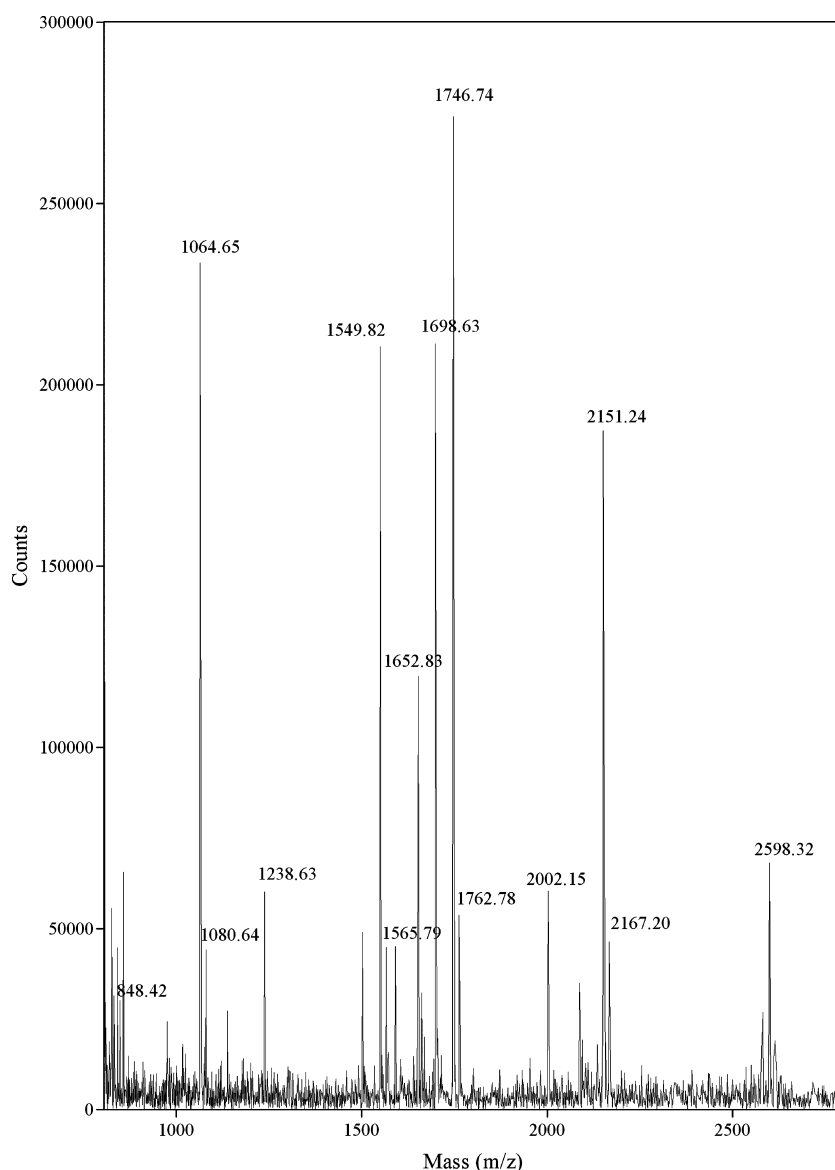


Fig. 1. MALDI-TOF mass spectrum acquired in reflector positive mode of peanut purified PSC33.

Table 1
Profound identification of PSC33

Measured mass (MH ⁺)	Computed mass	Error (ppm)	Start	End	Missed cuts	Peptide sequence
848.412	848.378	40	51	57	0	SSDQQQR
1064.642	1064.559	78	91	98	1	QMVQQFKR
1080.632	1080.554	72	91	98	1	QMVQQFKR (oxidized methionine)
1238.622	1238.510	91	6	16	1	GRQGDSSSCER
1549.812	1549.717	61	99	110	0	ELMNLPPQCCNFR
1565.787	1565.712	48	99	110	0	ELMNLPPQCCNFR (oxidized methionine)
1652.822	1652.828	−3	21	33	0	VNLKPCEQHIMQR
1746.732	1746.756	−14	34	47	0	IMGEQEYDSYDIR
1761.772	1761.751	12	34	47	0	IMGEQEYDSYDIR (oxidized methionine)
2002.142	2001.966	88	99	114	1	ELMNLPPQCCNFRAPQR
2151.232	2151.083	69	17	33	1	QVDRVNLKPCEQHIMQR
2167.192	2167.078	53	17	33	1	QVDRVNLKPCEQHIMQR (oxidized methionine)
2598.292	2598.090	78	71	90	1	CMCEALQIMENQCRLQDR

of protein with 40 µg of porcine trypsin (Promega) in 1.2 ml of 0.1 M Tris buffer, pH 8. The digestion solution was flushed with nitrogen to avoid oxidation of methionine and incubated overnight at 37 °C with gentle stirring. Peptides were separated on an analytical (250 × 4.6 mm) C18 reverse phase HPLC column (Vydac 218TP54) operated at 1 ml/min and 40 °C, and using the following gradient: 0–20% B in 15 min, 20–40% B in 120 min. Buffer A was 0.1% TFA in water, buffer B was 70% CH₃CN and 0.04% TFA in water. Fractions were collected at 30 s, checked for purity by MALDI-TOF in reflector positive mode and dried in a vacuum centrifuge. Pure fractions of the 3920 Da peak were resuspended in 50 µl of 70% TFA containing a few crystals of CNBr (Sigma-Aldrich). Hydrolysis was performed for 2 h at RT in the dark. The hydrolysate was diluted 10 times in CHCA matrix and analysed by MALDI-TOF in reflector negative mode.

Results

Proteomic analysis of natural PSC33

Several IgE immunoreactive fractions were obtained from peanut protein fractionation (H. Bernard, manuscript in preparation). One of them, purified to homogeneity, had an average MW of 14840 Da determined by MALDI-TOF and the common N-terminal sequences of PSC33 and Ara h 6. It gave one Coomassie blue-stained band in SDS-PAGE which was analysed by peptide mass fingerprinting after in-gel reduction, carbamidomethylation, and trypsin digestion (Fig. 1). The peptide mass list was used to search the whole NCBI nr protein database (all taxa, no MW and pI restriction) using the web accessible software Profound with a 150 ppm tolerance: the first hit, obtained with a probability of 1 was the *Arachis hypogaea* conglutin PSC33 with a 65% coverage of the protein sequence (Table 1). The second hit had only a

probability of 2.1×10^{-21} . It should be noted that the high intensity 1698.63 Da peak was not found as belonging to the PSC33 sequence which was the only one available in the database when the experiment was done (June 2003). However, this mass corresponded to that of the peptide 78–90 of a NEQ variant of PSC33 (the three polymorphism sites being contained in this peptide) whose sequence (AY722690) was recently (July 2004) deposited in the NCBI protein database. Thus two variants of PSC33 exist.

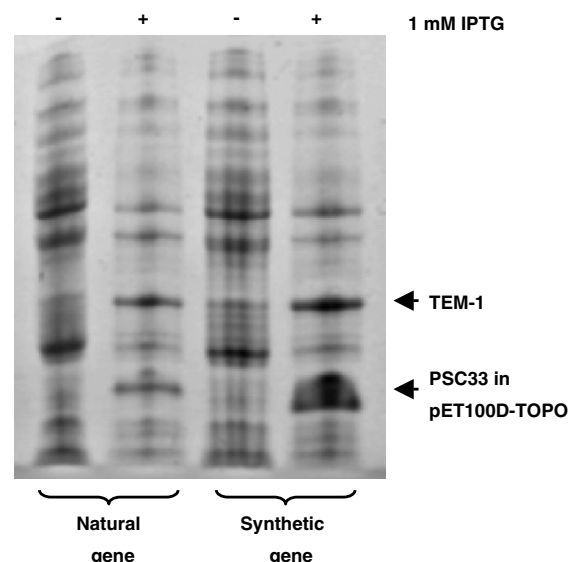


Fig. 2. Coomassie blue staining of SDS-PAGE on the Phast-system: comparison of PSC33 5-h expression in the same vector pET100D-TOPO and in the same bacteria BL21 (DE3) RIL when coded by the natural gene or the synthetic gene. The natural gene contains the propeptide (20 amino acids) whereas the synthetic gene does not: the mass of the natural gene product is thus slightly higher than that of the synthetic gene product. TEM-1 (vector β-lactamase) is located behind the inserted sequence on the expression vectors; it is co-expressed when its reading frame is in the same direction as that of PSC33.

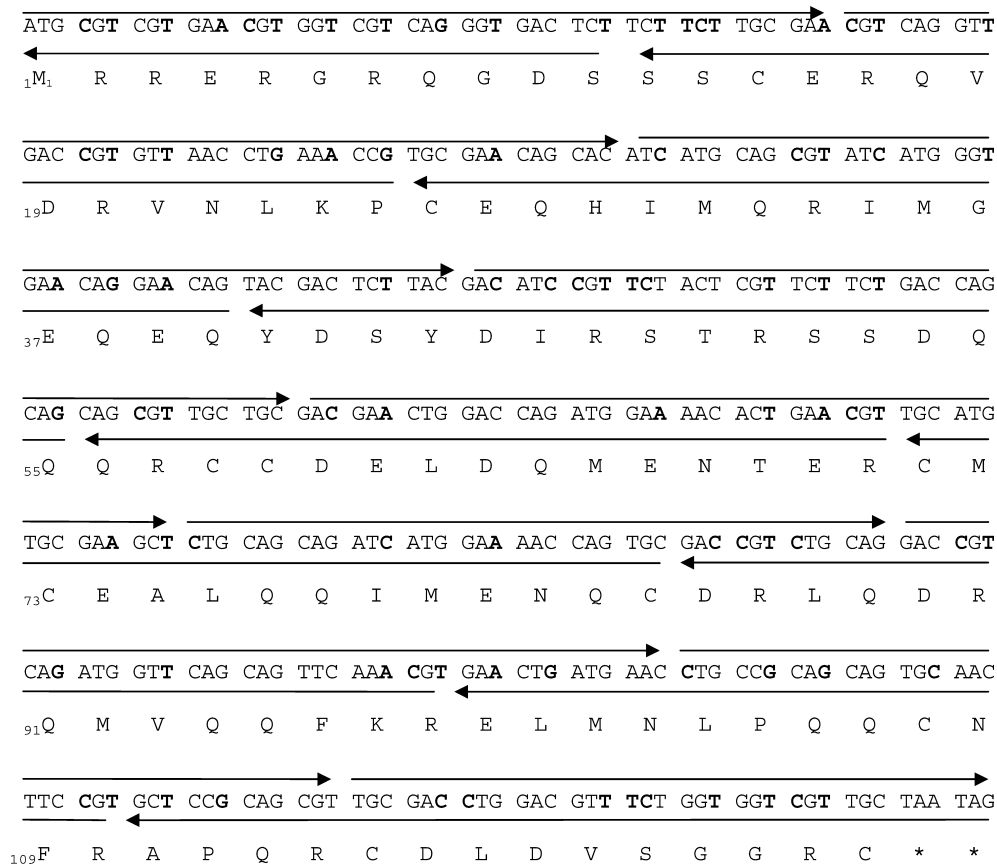


Fig. 3. Assembly of the synthetic oligonucleotides for the generation of the synthetic gene. Modified codons with respect to the natural gene are shown in bold. Arrows above and below the coding sequence represent respectively the forward and reverse oligonucleotides used for the assembly.

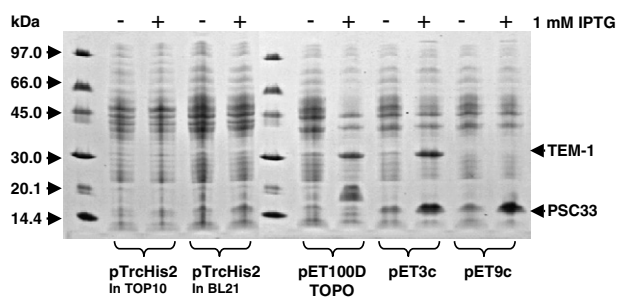


Fig. 4. Coomassie blue staining of SDS-PAGE on the Phast-system: 5-h expression of the PSC33 synthetic gene product in different vectors. In pET100D-TOPO, PSC33 contains 36 supplementary Tag amino acids in N-terminal.

Genomic cloning of PSC33

The PSC33 gene was amplified by PCR from peanut leaf genomic DNA. Using a forward primer starting at the mature protein N-terminal, 36 clones were sequenced in both directions: we observed that PSC33 was intron less

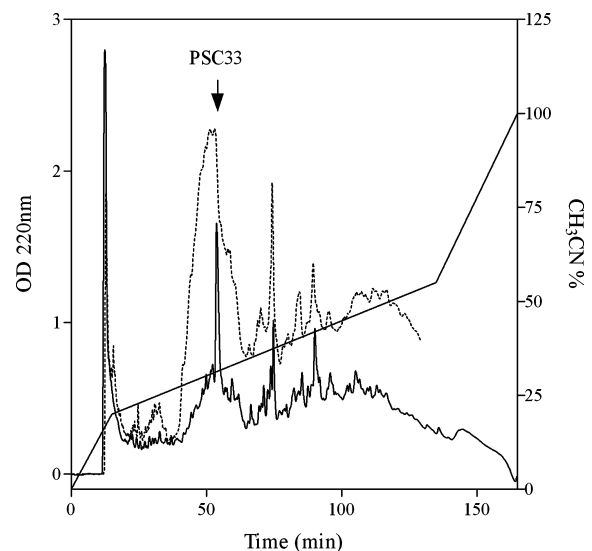


Fig. 5. Semi-preparative C18 reverse phase chromatography HPLC profile of refolded PSC33: solid line, refolding of a 200 ml culture in 10 ml; dotted line, refolding of a 500 ml culture in 40 ml.

and confirmed the existence of two variants of the PSC33 protein by obtaining 19 NEQ and 17 DQE variants.

Expression of the natural intron-less PSC33 gene

Since PSC33 is intron-less, it could be directly inserted into an expression vector and assayed for expression. There was no expression of this natural genomic PSC33 in the pTrcHis2-TOPO vector (whose His-tag is C-terminal) in TOP10 or BL21-Codonplus-RIL bacteria. One clone expressed the protein in the pTrcHis vector (His-tag in N-terminal) with no difference of expression level between TOP10 and BL21-Codonplus-RIL. Although we were able to purify the protein by its His-tag from this last clone, too little was recovered for refolding experiments. In the pET100D-TOPO vector natural PSC33 (with its propeptide) was much more expressed in BL21-Codonplus (DE3)-RIL than in BL21-(DE3) (not shown). However, this expression is still much less than when the synthetic gene of PSC33 (without its propeptide) is in the same vector and the same cells (Fig. 2).

Expression of the synthetic PSC33 gene

Considering these poor expressions and knowing that PSC33 has a high percentage of low usage codons in *E. coli* [18], we made a synthetic gene optimized for codon usage (Fig. 3). This synthetic gene was still poorly expressed when carried by the pTrcHis2-TOPO vector (Fig. 4, lanes 2–5). Expression was better with the pET100D-TOPO vector but β -lactamase TEM-1 was co-induced. The same phenomenon was observed in the pET3c vector: in these last two vectors, the host (PSC33) and ampicillin resistance (TEM-1) genes are in tandem and in the same direction. Since they are co-expressed, it is probably because the T7 terminator site (which is the same in pET-3c and pET100/D-TOPO vectors) is not efficient enough to stop T7 polymerase; the fact that PSC33 is a small sequence might also contribute to this effect. We shifted to the pET9c vector in which the kanamycin resistance gene is in the

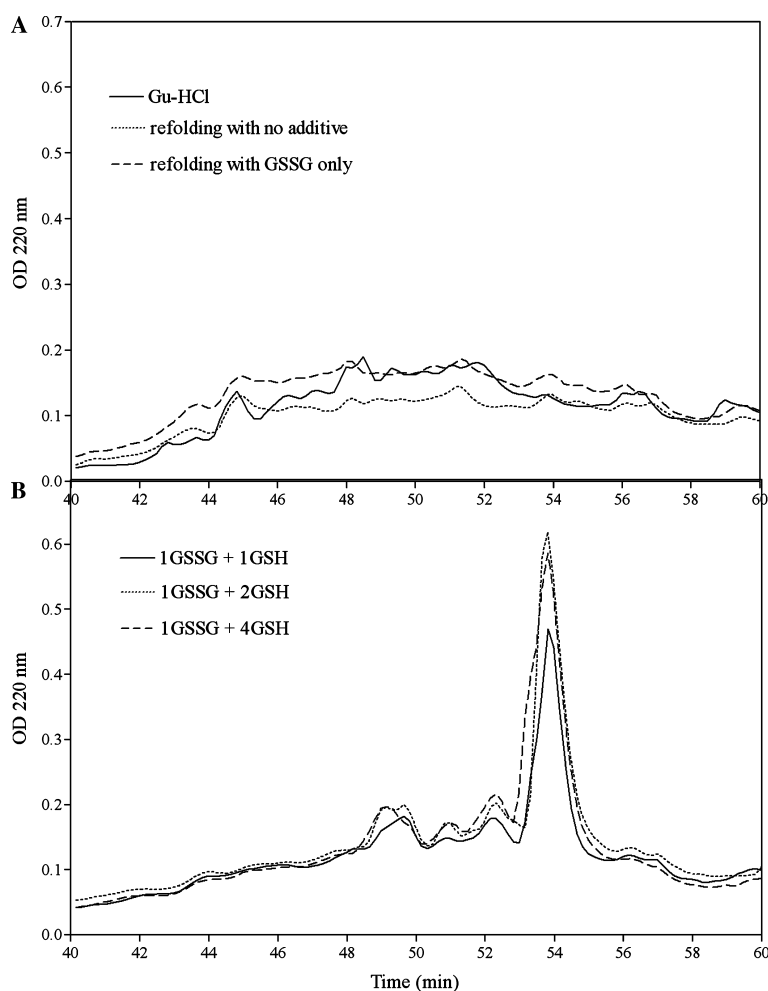


Fig. 6. Influence of oxidizing reagents on PSC33 refolding. Inclusion bodies obtained from 50 ml of bacterial culture were either directly submitted to HPLC after Gu-HCl solubilization or after refolding with oxidized (GSSG) and reduced (GSH) glutathione in various ratios. GSSG was always 250 μ M. Refolding volume was 4 ml and PSC33 was 17 μ M. Dialysate volume was 500 ml. (A) Conditions in which no refolding occurred. (B) Refolding occurrence.

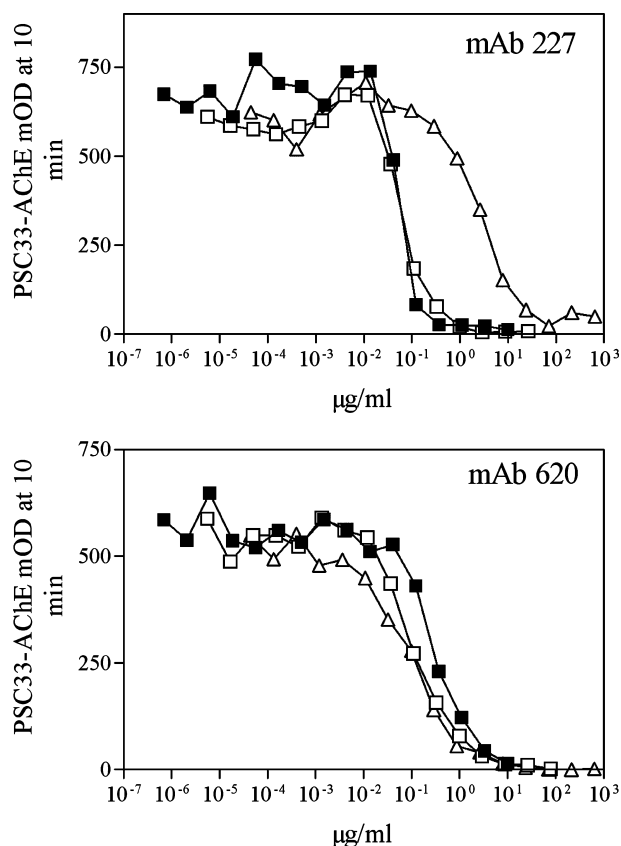


Fig. 7. Competitive enzyme-immunoassay of native and refolded PSC33. mAb 227 recognizes a conformational epitope, mAb 620 a linear epitope. (■), peanut purified PSC33; (□), refolded recombinant PSC33; and (△), non-refolded recombinant PSC33.

reverse direction and thus even if co-translated it is not co-expressed. In this last vector PSC33 is the main expression product.

Refolding and purification

These steps are described in Materials and methods. Since the proteins are expressed without tags, the whole protein content of the inclusion bodies was submitted to refolding. We checked on the tagged protein that Gu-HCl extracted the protein much better than urea. Purification by C18 reverse phase chromatography was performed after the refolding. Two semi-preparative C18 reverse phase HPLC profiles are shown in Fig. 5. PSC33 eluted at 54 min and at about 35% CH₃CN. In the first experiment, 1.1 mg of homogenous PSC33 was obtained from 200 ml of bacterial culture with an extraction and refolding volume of 10 ml. In the second trace, the inclusion bodies of 500 ml of culture were extracted and refolded in 40 ml: four fractions yielded a total of 5 mg of homogenous PSC33 with a proper immunoreactivity and disulfide bridge formation (see below). We studied the influence of reduced glutathione (GSH) by varying its concentra-

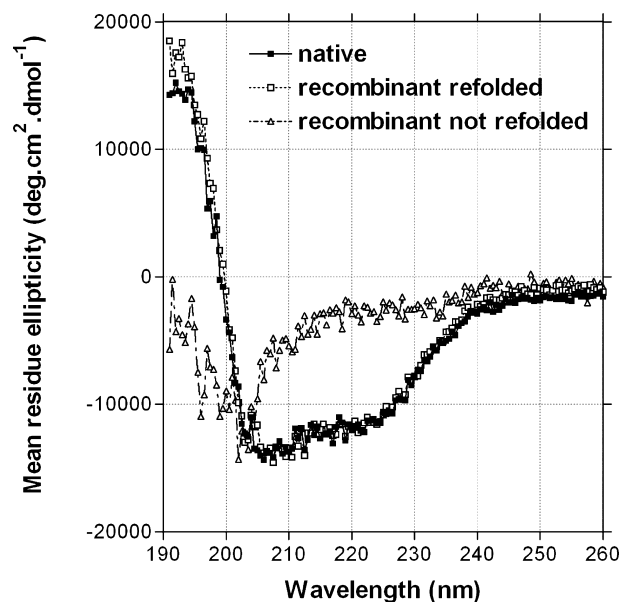


Fig. 8. Far-UV circular dichroism spectra of native (■), refolded (□), and non-refolded (△) PSC33. The value of θ in $\text{deg cm}^2 \text{dmol}^{-1}$ at 222 nm was used to calculate the α -helix content of each protein.

tion (Fig. 6). The absence of GSH resulted in the absence of a PSC33 peak as when the Gu-HCl extract was directly injected on the C18 column: in both cases PSC33 was detected by MALDI-TOF from 45 to 56 min and amounted to a total of 1 mg of total PSC33 when 50 ml of bacterial culture was extracted. The addition of 500 μM GSH gave the best refolding yield: 300 μg of lyophilized PSC33 was obtained from an extract of a 50 ml bacterial culture.

Assays of refolding: immunoreactivity and CD

Refolded and native PSC33 were compared in a competition immunoassay using mAbs raised against the native protein: two mAbs were used, one (mAb 620) that recognizes equally well the native and the reduced carbamidomethylated protein, i.e., a linear epitope and one (mAb 227) that recognizes only the native protein, and thus needs a conformational epitope to bind. Refolded and native PSC33 competed equally well with the acetylcholinesterase-labelled PSC33 for binding to mAb 227 whereas the non-refolded one bound 100-fold less (Fig. 7). The same result was obtained with three other mAbs that each recognize a different conformational epitope (not shown). In contrast and as expected, refolding had no effect on recognition of its linear epitope by mAb 620.

Native and refolded PSC33 showed the same far-UV CD spectrum (Fig. 8) with a 30% content in α -helix as estimated by the program K2d and the molar ellipticity at 222 nm [28,29]. For comparison, we checked the CD spectrum of the unfolded protein on the fraction eluting at 52 min in the Gu-HCl HPLC trace of Fig. 6A: it had only a 7% content in α -helix (Fig. 8).

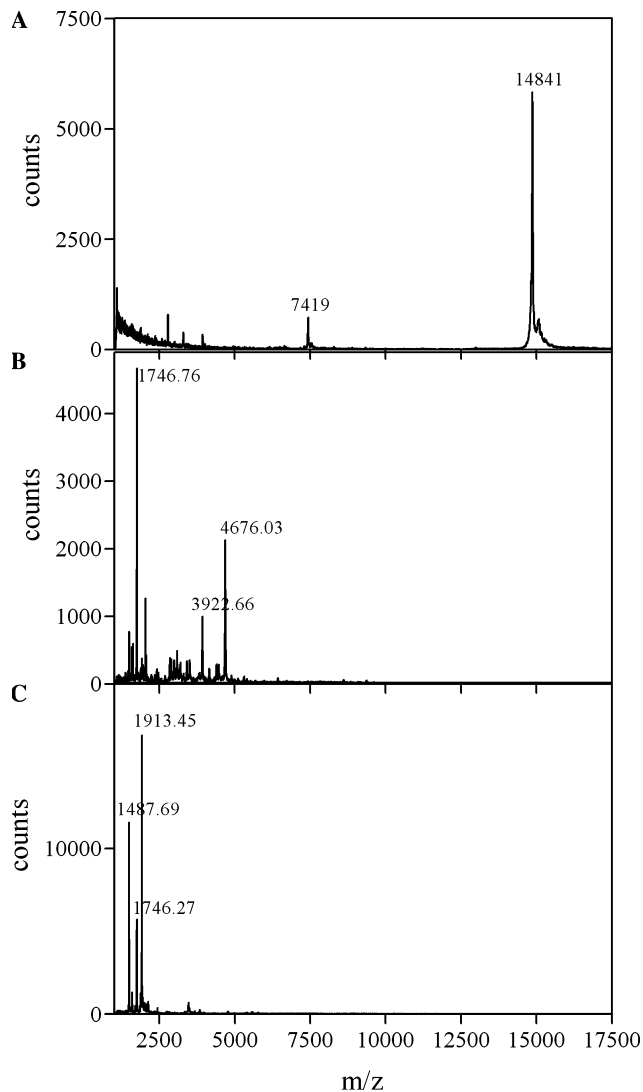


Fig. 9. MALDI-TOF mass spectrum acquired in linear positive mode of: (A) purified PSC 33; (B) the same sample digested overnight with 4% in weight porcine trypsin; and (C) the digestion product reduced with DTT. (Calibration was done on the undigested PSC33. Masses obtained in linear mode are not as accurate as those obtained in reflectron mode. The masses of the two unreduced tryptic peptides were also obtained more accurately using reflectron mode; not shown.)

Disulfide bridge identification

PSC33 differs from the other 2S albumins by two extra cysteines. Since the general pattern of 2S albumin disulfides is known it was easy to draw a model of the putative extra bridge of PSC33. After overnight trypsin digestion two main DTT-sensitive peptides were identified by MALDI-TOF (Fig. 9). Their MW (3921 and 4372 Da) gave disulfide linked peptides coherent with the general 2S albumin pattern (Figs. 10A and B). These two peaks were C18 purified, reduced with DTT and rechromatographed to confirm their composition (data not shown).

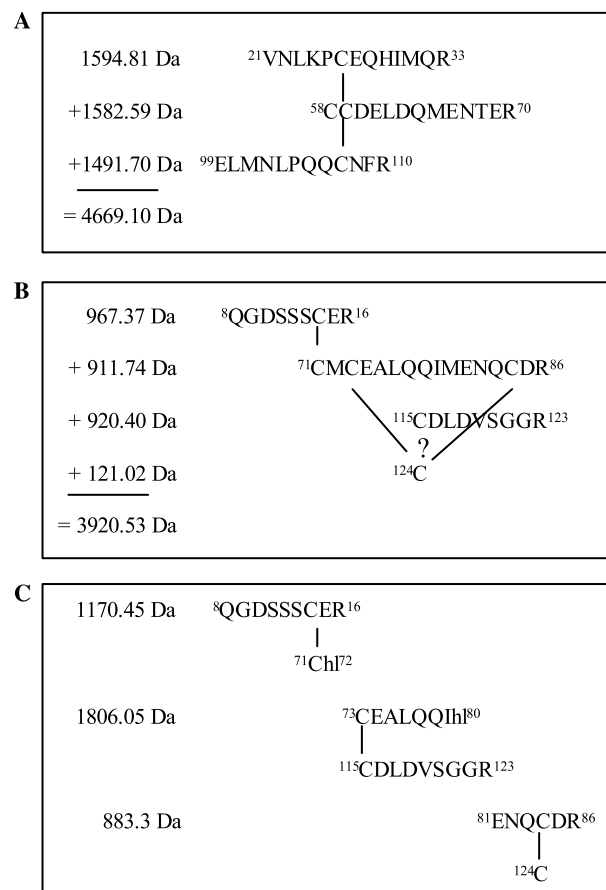


Fig. 10. Disulfide arrangement in PSC33. (A, B) Models of the plausible disulfide arrangement of PSC33 deduced from the mass of the two unreduced tryptic peptides seen on the MALDI-TOF mass spectrum of Fig. 9B and from the known disulfide arrangement of other 2S albumins. (A) Arrangement of the 4676 Da peak. (B) Arrangement of the 3920 Da peak. (C) $M + H^+$ of the peptides obtained after CNBr digestion of the 3920 Da peak. hl, homoserine lactone.

Since the two cysteines of the 4372 Da peak are contiguous, no further position assignment was feasible. On the other hand, each cysteine of the 3920 Da peak was separated from the other by a methionine and a further cleavage allowed resolution of the ambiguities in the pairing of cysteines 73 and 84 (Fig. 10). MALDI-TOF analysis in the reflector negative mode of the CNBr cleavage of the 3920 Da peak yielded three peaks of MH^+ 883.3, 1170.45, and 1806.05 Da (these last two masses are those of the homoserine lactone form of the methionine) corresponding, respectively, to the pairing of C84 to C124, C14 to C71, and C73 to C115 (Fig. 10C). The final model of disulfide pairing of PSC33 is depicted in Fig. 11. The same results were obtained with native PSC33.

Discussion

2S albumins are now clearly identified as legumes pan-allergens. The fact that we identified in peanut the

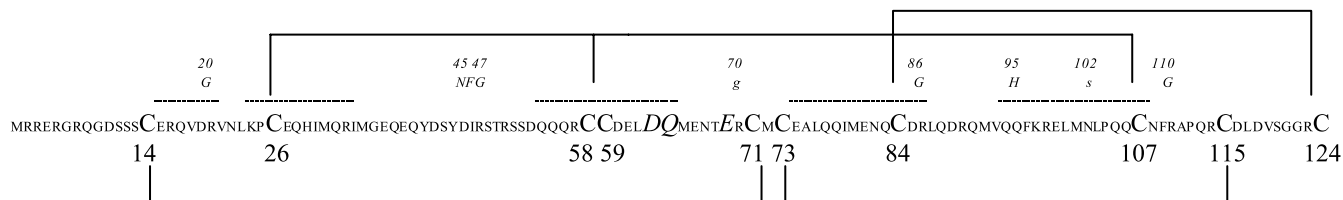


Fig. 11. Final model of the natural and refolded PSC33 disulfide arrangement. Cysteines are in bold upper case. The three polymorphic residues are in bold italic upper case. The residues found mutated in the original Ara h 6 (AF092846) and a new Ara h 6 (AY871100) are above the PSC33 (AF366561) sequence in light italic upper case and lower case, respectively. Dotted lines represent the five α -helices determined in Ara h 6 by NMR [9].

previously described PSC33 protein and not the highly homologous Ara h 6 [21,22] might be due to the known polymorphism of these proteins [1]. We illustrated this polymorphism by the genomic cloning of two of these variants whose sequences have already been deposited in protein databases.

A prerequisite before attempting to refold a protein from inclusion bodies is to have it in sufficient amounts. A synthetic gene of PSC33 was made to circumvent the great difference in codon usage between *Arachis hypogaea* and *E. coli*. This synthetic gene was only optimized for codon usage and not for other parameters such as presence of internal ribosome entry sites, mRNA secondary structures or direct DNA repeats. However, the synthetic PSC33 gene was still very poorly expressed in vectors using *E. coli* RNA polymerase under Trc promoter control (pTrc vectors) and we had to shift to pET vectors using the T7 RNA polymerase to obtain a proper expression. Providing *E. coli* with rare codons tRNA (BL21-Codonplus (DE3)-RIL) [18] increased the expression of the natural gene but the level of expression was still much lower than that of the synthetic gene in conventional bacteria. Until now, synthetic genes of three allergens have been made: Fel d 1, Der p 1, and bee venom phospholipase A2.

As stated in Introduction, recombinant 2S albumins have been produced by several ways but not by in vitro refolding from inclusion bodies. In contrast, several reports have been published on refolding of proteins from the neighbour subfamilies: wheat amylase inhibitors (WAI) CM16 and 0.19 [15,16], corn Hageman factor inhibitor [23], and the wheat non-specific lipid transfer protein 1 [24]. We had successfully refolded in the laboratory a scFv using a method developed by Tsumoto et al. and called 'additive-introduced stepwise dialysis' [17,25]. ScFv are β -sheet rich proteins, and the same authors modified the method to refold an α -helix rich protein (IL-21) [26] by adding as oxidizing agent not only GSSG but a mixture of GSSG and GSH with GSH in excess. Our results confirm this finding, showing that protein secondary structure influence their disulfide exchange. The yield of native PSC33 was 0.6 mg for 100 ml of bacterial culture when up to 200 ml of culture was treated. However when processing 500 ml of culture, the yield increased to 1 mg per 100 ml of culture (Fig. 5).

In addition to CD and immunoreactivity studies of refolding, we also used tryptic and CNBr digestions to identify the peptides linked by disulfide bridges by MALDI-TOF mass spectrometry. We showed that these links were the same in the natural and refolded proteins. PSC33 has two more cysteines than the other 2S albumins. The first one, C84, is also found at the same position in lupin conglutin δ and WAI 0.19, 0.28, and 0.53 [27]. In conglutin δ it is unpaired whereas in WAIs it is paired with C28 and C29. In PSC33, this C84 is paired with the second extra-cysteine C124 which is C-terminal. This fifth disulfide bridge thus involves the two supplementary cysteines and conserves the canonical disulfide pattern of 2S albumins. PSC33 should thus be more stable to heat and proteolysis than the 4-disulfides 2S albumins such as Ara h 2. Removal of this fifth disulfide bridge should permit us to test this hypothesis.

Acknowledgments

The authors thank Sandrine AH-LEUNG for bringing back the fresh peanut seeds from Réunion island and Marie Françoise DRUMARE for assistance in preparation of monoclonal antibodies.

References

- [1] R.I. Monsalve, M. Villalba, M. Rico, P.R. Shewry, R. Rodriguez, The 2S albumin proteins, in: E.N.C. Mills, P.W. Shewry (Eds.), Plant Food Allergens, Blackwell, Oxford, 2003, pp. 42–56.
- [2] C.A. Behnke, V.C. Yee, I. LeTrong, L.C. Pedersen, R.E. Stenkamp, S.S. Kim, G.R. Reeck, D.C. Teller, Structural determinants of the bifunctional corn hageman factor inhibitor: X-ray crystal structure at 1.95 Å resolution, *Biochemistry* 37 (1998) 15277–15288.
- [3] J.A. Jenkins, S. Griffiths-Jones, P.R. Shewry, H. Breiteneder, E.N.C. Mills, Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis, *J. Allergy Clin. Immunol.* 115 (2005) 163–170.
- [4] D. Mittag, J. Acedias, B.K. Ballmer-Weber, L. Vogel, M. Wensing, W.M. Becker, S.J. Koppelman, A.C. Knulst, A. Hebling, S.L. Hefle, R. van Ree, S. Vieths, Ara h 8, a Bet v 1-homologous allergen from peanut, is a major allergen in patients with combined birch pollen and peanut allergy, *J. Allergy Clin. Immunol.* 114 (2004) 1410–1417.

- [5] O.G. Paik-Ro, J.C. Seib, R.L. Smith, Seed-specific, developmentally regulated genes of peanut, *Theor. Appl. Genet.* 104 (2002) 236–240.
- [6] D. Pantoja-Uceda, O. Palomares, M. Bruix, M. Villalba, R. Rodriguez, M. Rico, J. Santoro, Solution structure and stability against digestion of rproBnIb, a recombinant 2S albumin from rapeseed: relationship to its allergenic properties, *Biochemistry* 43 (2004) 16036–16045.
- [7] D. Pantoja-Uceda, P.R. Shewry, M. Bruix, A.S. Tatham, J. Santoro, M. Rico, Solution structure of a methionine-rich 2S albumin from sunflower seeds: relationship to its allergenic and emulsifying properties, *Biochemistry* 43 (2004) 6976–6986.
- [8] D. Pantoja-Uceda, M. Bruix, G. Gimenez-Gallego, M. Rico, J. Santoro, Solution structure of RicC3, a 2S albumin storage protein from *Ricinus communis*, *Biochemistry* 42 (2003) 13839–13847.
- [9] K. Lehmann, K. Schweimer, P. Neudecker, P. Rösch, Sequence-specific 1H, 13C and 15N resonance assignments of Ara h 6, an allergenic 2S albumin from peanut, *J. Biomol. NMR* 29 (2004) 93–94.
- [10] G. Clement, D. Boquet, Y. Frobert, H. Bernard, L. Negroni, J.M. Chatel, K. Adel-Patient, C. Creminon, J.M. Wal, J. Grassi, Epitopic characterization of native bovine β -lactoglobulin, *J. Immunol. Methods* 266 (2002) 67–78.
- [11] M.J.C. Alcocer, G.J. Murtagh, P.B. Wilson, P. Progijs, J. Lin, D.B. Archer, The major human structural IgE epitope of the brazil nut allergen Ber e 1: a chimaeric and protein microarray approach, *J. Mol. Biol.* 343 (2004) 759–769.
- [12] M.A. Gonzalez de lapena, R.I. Monsalve, E. Batanero, M. Villalba, R. Rodriguez, Expression in *Escherichia coli* of Sin a 1 the major allergen from mustard, *Eur. J. Biochem.* 237 (1996) 827–832.
- [13] K. Lehmann, S. Hoffmann, P. Neudecker, M. Suhr, W.M. Becker, P. Rösch, High-yield expression in *Escherichia coli*, purification, and characterization of properly folded major peanut allergen Ara h 2, *Protein Expr. Purif.* 31 (2003) 250–259.
- [14] C. Fernandez-Tornero, A. Ramon, M.L. Navarro, J. Varela, G. Gimenez-Gallego, Synthesis of proteins with disulfide bonds in *E. coli* using defined culture media, *Biotechniques* 32 (2002) 1238–1242.
- [15] V. Lullien-Pellerin, S. Gavalda, P. Joudrier, M.F. Gautier, Expression of a cDNA encoding the wheat CM16 protein in *Escherichia coli*, *Protein Expr. Purif.* 5 (1994) 218–224.
- [16] M. Okuda, T. Satoh, N. Sakurai, K. Shibuya, H. Kaji, T. Samejima, Overexpression in *Escherichia coli* of chemically synthesized gene for active 0.19 α -amylase inhibitor from wheat kernel, *J. Biochem.* 122 (1997) 918–926.
- [17] K. Tsumoto, K. Shinoki, H. Kondo, M. Uchikawa, T. Juji, I. Kumagai, Highly efficient recovery of functional single-chain Fv fragments from inclusion bodies overexpressed in *Escherichia coli* by controlled introduction of oxidizing reagent- application to a human single-chain Fv fragment, *J. Immunol. Methods* 219 (1998) 119–129.
- [18] T. Kleber-Janke, W.M. Becker, Use of modified BL21(DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage, *Protein Expr. Purif.* 19 (2000) 419–424.
- [19] J. Havlis, H. Thomas, M. Sebela, A. Shevchenko, Fast-response proteomics by accelerated in-gel digestion of proteins, *Anal. Chem.* 75 (2003) 1300–1306.
- [20] L. Negroni, H. Bernard, G. Clement, J.M. Chatel, P. Brune, Y. Frobert, J.M. Wal, J. Grassi, Two-site enzyme immunometric assays for determination of native and denatured β -lactoglobulin, *J. Immunol. Methods* 220 (1998) 25–37.
- [21] T. Kleber-Janke, R. Crameri, U. Apenzeller, M. Schlaak, W.M. Becker, Selective cloning of peanut allergens including profilin and 2S albumins, by phage display technology, *Int. Arch. Allergy Immunol.* 119 (1999) 265–274.
- [22] M. Suhr, D. Wicklein, U. Lepp, W.M. Becker, Isolation and characterization of natural Ara h 6: evidence for a further peanut allergen with putative clinical relevance based on resistance to pepsin digestion and heat, *Mol. Nutr. Food Res.* 48 (2004) 390–399.
- [23] M. Hazegh-Azam, S.S. Kim, S. Masoud, L. Andersson, F. White, L. Johnson, S. Muthukrishnan, G. Reeck, The corn inhibitor of activated hageman factor: purification and properties of two recombinant forms of the protein, *Protein Expr. Purif.* 13 (1998) 143–149.
- [24] V. Lullien-Pellerin, C. Devaux, T. Ithorai, D. Marion, V. Pahin, P. Joudrier, M.F. Gautier, Production in *Escherichia coli* and site-directed mutagenesis of a 9-kDa nonspecific lipid transfer protein from wheat, *Eur. J. Biochem.* 260 (1999) 861–868.
- [25] K. Makabe, R. Asano, T. Ito, K. Tsumoto, T. Kudo, I. Kumagai, Tumor-directed lymphocyte-activating cytokines: refolding-based preparation of recombinant human interleukin-12 and an antibody variable domain-fused protein by additive-introduced stepwise dialysis, *Biochem. Biophys. Res. Commun.* 328 (2005) 98–105.
- [26] R. Asano, T. Kudo, K. Makabe, K. Tsumoto, I. Kumagai, Antitumor activity of interleukin-21 prepared by novel refolding procedure from inclusion bodies expressed in *Escherichia coli*, *FEBS Lett.* 528 (2002) 70–76.
- [27] T.A. Egorov, T.I. Odintsova, A.K. Musolyamov, R. Fido, A.S. Tatham, P.R. Shewry, Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily, *FEBS Lett.* 396 (1996) 285–288.
- [28] M.J. Pandya, R.B. Sessions, P.B. Williams, C.E. Dempsey, A.S. Tatham, P.R. Shewry, A.R. Clarke, Structural characterization of a methionine-rich, emulsifying protein from sunflower seed, *Proteins* 38 (2000) 341–349.
- [29] C.A. Rohl, R.L. Baldwin, Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides, *Biochemistry* 36 (1997) 8435–8442.