

Mohamed Elfatih H. Bashir^a
Ina Hubatsch^a
Hans Peter Leinenbach^a
Michael Zeppezauer^a
Raphael C. Panzani^b
Ismail H. Hussein^c

^a University of Saarland, Saarbrücken, Germany;

^b Laboratoire de Recherches, Marseille, France;

^c University of Gezira, National Oilseed Processing Research Institute (NOPRI), Wad Medani, Sudan

Ric c 1 and Ric c 3, the Allergenic 2S Albumin Storage Proteins of *Ricinus communis*: Complete Primary Structures and Phylogenetic Relationships

Abstract

The 2S albumin storage protein of *Ricinus communis* consists of the two heterodimeric proteins Ric c 1 and Ric c 3 each of which is composed of a small and a large subunit linked together by disulphide bridges. The complete primary structures of both heterodimeric proteins were determined by enzymatic degradation and automated Edman degradation. The sequences of all four chains correspond to the known cDNA sequence of the gene of a presumed precursor molecule and to the previously determined partial sequences for Ric c 1 and Ric c 3. In addition, few differences in amino acid positions were found which seem to be related to different varieties of *R. communis*. Sequence comparisons with 2S albumin from other plant genera revealed high degrees of homology and support the view of a common genetic origin of this protein family. Ric c 1 and Ric c 3 which have 11,212 and 12,032 daltons, respectively, share a similar molecular size, biological function and allergenicity with the 2S albumins from *Brassica juncea* (Bra j 1E) and *Sinapis alba* L (Sin a 1). Ric c 1 and Ric c 3 may be classified as isoallergens if, additionally, the high degree of similarity in the position of polar residues is taken into account.

Key Words

Ric c 1

Ric c 3

Ricinus communis allergens

Castor bean

Heterodimer

2S albumin storage proteins

Amino acid sequence

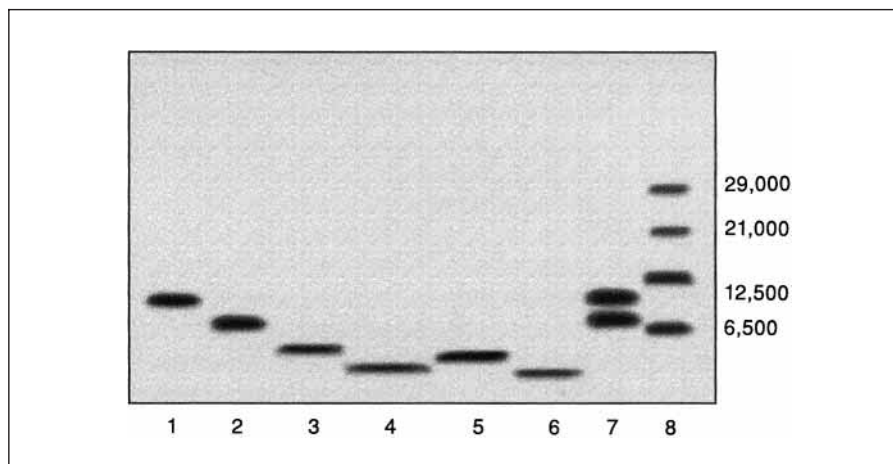
Introduction

Castor bean allergens have been the subject of extensive studies for the last 5 decades due to the high allergenic potential of this plant species [1–4]. Previous studies on the aqueous extract of *Ricinus communis* revealed a multiplicity of antigenic and allergenic components with a wide range of molecular sizes [1, 5–8]. Among the water soluble proteins the 2S albumin storage protein (2S ASP) is of particular interest since Youle and Huang [9] first described their allergenic nature. Thorpe et al. [5] found that most (96%) castor bean-sensitive patients tested had specific IgE antibody to the purified 2S storage albumin, confirming it as the major allergen of castor bean. Sharief and Li [10] de-

termined the amino acid sequence of the so-called glutamine-rich seed storage protein, and showed that it consisted of small and large subunits covalently linked by disulphide bonds.

Irwin et al. [11, 12] using probes based on the sequence of Sharief and Li [10] isolated a cDNA coding for a 34-kDa precursor of the *R. communis* 2S albumin. They suggested from the cDNA sequence analysis that a single preproprotein may be processed into two different heterodimeric storage proteins [11, 12]. One of these two different 2S ASP, as expected, corresponds to the small and large subunits sequenced by Sharief and Li [10], which correspond to the regions 136–169 and 173–237 in the genome of the precursor protein [11, 12]. The other putative protein encoded with

Fig. 1. PAGE of Ric c 1 and Ric c 3 and their constituent peptide chains. Different protein fractions were analyzed by electrophoresis in 16% SDS-PAGE: lane 1: 10 μ g Ric c 3; lane 2: 10 μ g Ric c 1, lane 3: large subunit Ric c 3 (A β), lane 4: small subunit Ric c 3 (A α), lane 5: large subunit Ric c 1 (B β), lane 6: small subunit Ric c 1 (B α), lane 7: 2S ASP non-reduced, and lane 8: 10 μ g molecular weight standards.



the same gene was proposed by Irwin et al. [11, 12] to occur as a further component of the 2S ASP. Recently, Machado and Silva [13], applying the purification scheme of Spies and Coulson [14], have aligned a putative allergen, on the basis of amino acid composition and partial amino acid sequence data with the cDNA sequence of Irwin et al. [11, 12]. Based on the amino acid composition and N-terminal sequencing of Ric c 3 allergen (previously called CB-1A [14]), sequence identity with residues 18–50 and 66–130 of the cDNA has been revealed, supporting the view that Ric c 3 is the putative 2S ASP encoded with the same gene.

In order to confirm the presumed relationships between cDNA and the corresponding allergens in *R. communis*, the complete primary structure of both heterodimeric allergens 2S ASP code A (Ric c 3) and 2S ASP code B (Ric c 1) has been determined. The establishment of the complete amino acid sequence of Ric c 1 and Ric c 3 should be important for future work concerning the precise location of B and T cell epitopes on the allergen and provide a further step in establishing structure-function correlations in this particular field of immediate hypersensitivity. Both Ric c 1 and Ric c 3 were purified to homogeneity and examined by ELISA against sera from castor bean-sensitive patients. Specific IgE and IgG antibodies were detected in all samples, confirming both proteins as the major allergens. In spite of a 100% frequency of IgE binding to both Ric c 1 and Ric c 3, the immunological data reveal certain differences between the individual sera as to their IgE and IgG binding capacity and specificity [15].

We have confirmed that each of the heterodimeric allergens Ric c 1 and Ric c 3 containing 99 and 104 residues, respectively, are composed of covalently S-S-linked small and large subunits. Both large subunits are N-terminally

blocked. In addition we have determined the complete primary structure of each heterodimer. The molecular masses of Ric c 3 and Ric c 1 are 12,032 and 11,212 daltons, respectively. The sequence analysis described below indeed proves that both species derive from the common precursor protein identified at the gene level by Irwin et al. [11, 12], and reveals interesting phylogenetic relationships to other allergenic 2S albumins from different plant genera.

Materials and Methods

Extraction of Castor Beans

A sample of raw castor beans, Pacific-6 variety, was obtained from the Agriculture Research Corporation, Wad Medani, Sudan. The castor beans were extracted with PBS according to the method of Lehrer et al. [7, 8]. Gel filtration was performed on a Sephadex G-50 superfine column (2.5×90 cm) equilibrated by PBS. 100–200 mg of the lyophilized sample were applied to the column and 5 ml fractions were collected. Fractions were analyzed for protein by ultraviolet (UV) absorption at 280 nm. Elution with PBS as eluent revealed four main peaks designated chronologically V, F₁, F₂ and F₃.

Purification of 2S ASP

The method described by Thorpe et al. [5] was modified and performed for the preparation and purification of 2S albumins. After separation of the ricin (fraction V), the isolated fraction F₁ was used for purification of the storage albumins. The protein was precipitated with 90% saturated ammonium sulphate at 4°C overnight, and the precipitated protein was centrifuged at 2,000 *g* for 15 min. The precipitate was dissolved in 5 ml 0.05 mol/l of imidazole HCl buffer, pH 7.0, and run on a Sephadex G-50 superfine column (2.5×90 cm). The eluted protein in the first fraction was reprecipitated in 90% ammonium sulphate as before and rechromatographed on Sephadex G-50 to produce a pure preparation of the 2S albumins.

Reverse-Phase High-Performance Liquid Chromatography

The 2S ASP from fraction F₁, prepared by repeated gel filtration, were fractionated by high performance liquid chromatography (HPLC) on a C₁₈ column 100 Å 5 µm (Nucleosil ET 250/8/4, 250×4 mm, Macherey & Nagel, Düren, Germany) at a flow rate of 1 ml/min using a Waters 600 gradient HPLC system (Multisolute Delivery System, Waters-Millipore, Eschborn, Germany). UV absorption was recorded at 214 nm. The solvent system consisted of 0.1% trifluoroacetic acid (TFA) in H₂O (v/v) as eluent A and 60% acetonitrile in 0.1% TFA (v/v) as eluent B. 100 µl (250 nmol) of a 2S ASP solution were injected and eluted by a linear gradient (0–100% eluent B) for 60 min. The two components of the 2S ASP eluted at 36 and 39 min. The collected fractions were concentrated by freeze-drying and re-chromatographed under the same conditions. After lyophilization the fractions were stored at –25°C.

Allergen Nomenclature

We have adopted the nomenclature for individual components of castor bean allergen which is codified in the IUIS Allergen Nomenclature System [16, 17]. HPLC chromatography of fraction F₁, i.e. 2S ASP, showed two symmetrical peaks (see above). The protein eluting with a retention time of 36 min corresponding to the 2S ASP code A should be named Ric c 3, the second one with a retention time of 39 min corresponding to the 2S ASP code B should be named Ric c 1.

Reductive Cleavage and S-Alkylation of Ric c 3 and Ric c 1 with 4-Vinylpyridine

100 µg of the individual purified allergens Ric c 3 (code A) and Ric c 1 (code B) were dissolved in 150 µl 25 mM Tris-HCl buffer pH 8.2, containing 6 M guanidinium-HCl, and 0.1 mM EDTA. Reduction was carried out in the presence of dithiothreitol (DTT). DTT was added to a final concentration of 1 mM to reduce the interchain disulphide bonds by vortexing and incubating at 37°C for 2 h until no obvious pellet formed on centrifugation. A stock solution of 4-vinylpyridine in water was prepared and dissolved in water to the final concentration of 3.5 mM. Alkylation was accomplished by the addition of 5 µl 3.5 mM 4-vinylpyridine and the reaction mixture was incubated for 2 h at 37°C in the dark. Reduced alkylated proteins were freed from the reduction and alkylation reagents by reverse-phase HPLC (RP-HPLC). The solvent system consisted of 0.1% TFA in H₂O (v/v) as eluent A and 60% acetonitrile in 0.1% TFA (v/v) as eluent B. The samples were injected on RP-HPLC and eluted 15 min isocratically (3% eluent B) and then with a linear gradient (3–100% eluent B) for 60 min. One-milliliter fractions were collected, and 100-µl aliquots of each fraction were freed from acetonitrile at 35°C. Finally the fractions containing the disulphide-reduced and S-alkylated constituent peptide chains Aα, Aβ, Bα and Bβ were separated after rechromatography by RP-HPLC, pooled and lyophilized. The pyridylethylated peptides Aα and Bα were sequenced directly whereas blocked Aβ and Bβ peptides were sequenced after enzymatic cleavage.

Enzymatic Cleavage

Tryptic Digestion and Cleavage with Staphylococcal Endoproteinase Glu-C (V8 Protease). The isolated, N-terminally blocked, large subunits Aβ and Bβ (100 µg) were cleaved with TPCK-treated trypsin from bovine pancreas and staphylococcal endoproteinase Glu-C (V8 protease) according to Aitken et al. [18].

Digestion with Pyroglutamate Aminopeptidase. The isolated, N-terminally blocked, large subunits Aβ and Bβ (50 µg) were digested in 50 mM (NaH₂PO₄/Na₂HPO₄)/5 mM EDTA/5 mM DTT, pH 8, with

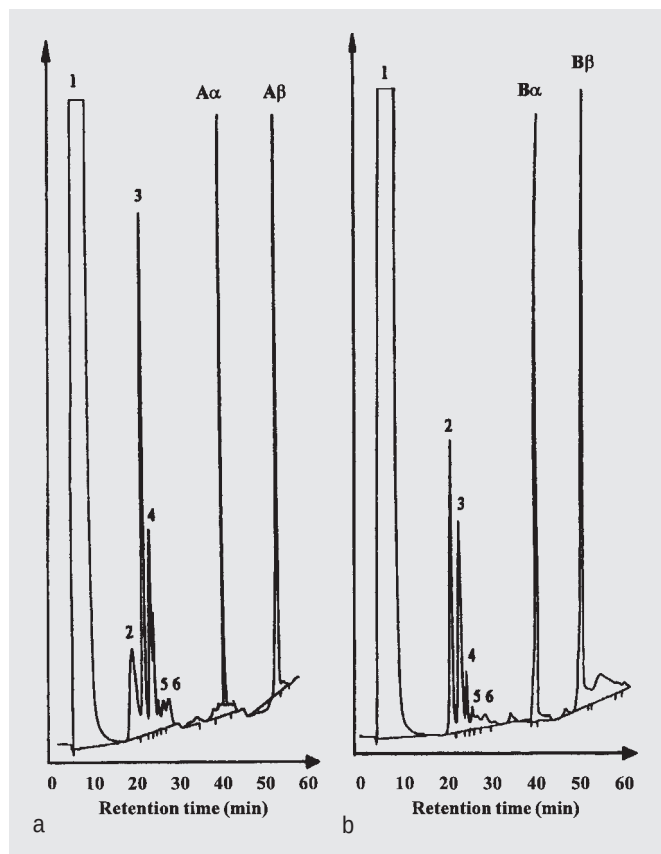


Fig. 2. Reduction, derivatization and RP-HPLC separation of the subunits of Ric c 1 and Ric c 3. Experimental conditions: column: NUCLEOSIL C₁₈ 100 Å 5 µm, 250×4 mm, ID. Injection volume: 50 µl (50 µg); eluent A: TFA 0.1% (v/v) in H₂O; eluent B: TFA 0.1% (v/v) in water, acetonitrile 60% (v/v), eluent: 15 min isocratic 3% B, then linear gradient from 3 to 100% B in 60 min, flow rate 1 ml/min; detection 214 nm. a HPLC profile of the Ric c 3 (2 subunits Aα and Aβ). b HPLC profile of the Ric c 1 (2 subunits Bα and Bβ). Peaks marked 1–6 were also obtained when the derivatization procedure was applied without protein. Numbered peptides Aα and Bα were submitted to amino acid sequencing directly whereas blocked Aβ and Bβ were sequenced after enzymatic cleavage.

pyroglutamate aminopeptidase from calf liver (EC 3.4.19.3). The enzyme-to-substrate ratio was 1:50, and the reaction was carried out at 37°C for 24 h in the presence of 4 M urea.

N-Terminal Amino Acid Sequencing

After RP-HPLC the collected fractions were freed from acetonitrile at 35°C and stored at –25°C. An aliquot of the sample containing about 200 pmol peptide was spotted onto a glass fibre membrane and dried with a stream of nitrogen. Sequence analyses was performed on an Applied Biosystems 473A peptide sequencer containing a Micro Cartridge (Applied Biosystems, Weiterstadt, Germany). Sequential grade reagents were used for automated Edman degradation (Applied Biosystems).

Fig. 3. Alignment of the amino acid sequences of Ric c 3 and Ric c 1 with the corresponding cDNA sequence and the partial sequences reported previously. The cDNA sequence is from Irwin et al. [12], the 2S ASP sequence is from Sharief and Li [10] and the sequences of Ric c 3 and Ric c 1 are from this study. The data of Machado and Silva [13] contain the amino acid sequence of six tryptic peptides determined by Edman degradation. The 2S ASP previously sequenced by Sharief and Li [10] had 94% similarity to Ric c 1 prepared in this study. Hyphenated gaps indicate amino acids which are missing in the sequences reported by Machado and Silva [13] or Sharief and Li [10].

10	20	30	cDNA sequence
S F A Y R T T I T T I E I D E S K G E R E G S S S Q Q C R Q	E S K G E R E G S S S Q Q C R Q		Ric c 3 small subunit A α
	G E - E - - - - -		data of Machado and Silva, 1992
40	50	60	cDNA sequence
E V Q R K D L S S C E R Y L R Q S S S R R S P G E E V L R M			Ric c 3 small subunit A α
- - - - -	- - - - -	- - - - -	data of Machado and Silva, 1992
70	80	90	cDNA sequence
P G D E N Q Q Q E S Q Q L Q Q C C N Q V K Q V R D E C Q C E			Ric c 3 large subunit A β
	Q Q Q E S Q Q L Q Q C C N Q V K Q V R D E C Q C E		data of Machado and Silva, 1992
	- - - - -	- - - - -	
100	110	120	cDNA sequence
A I K Y I A E D Q I Q Q G Q L H G E E S E R V A Q R A G E I			Ric c 3 large subunit A β
A I K Y I A E D Q I Q Q G Q L H G E E S E R V A Q R A G E I			data of Machado and Silva, 1992
- - - Y I A E D Q I Q Q - - L H G E E - - - V A Q R A G E I			
130	140	150	cDNA sequence
V S S C G V R C M R Q T R T N P S Q Q G C R G Q I Q E Q Q N			Ric c 3 large subunit A β and
V S S C G V R C M R Q T	P S Q Q G C R G Q I Q E Q Q N		Ric c 1 small subunit B α
V - - - - R C M - - -	P S Q Q G C R G Q I Q E Q Q N		data of Machado and Silva, 1992
			2S ASP of Sharief and Li, 1982
160	170	180	cDNA sequence
L R Q C Q E Y I K Q Q V S G Q G P R R S D N Q E R S L R G C			Ric c 1 small subunit B α and
L R Q C Q E Y I K Q Q V S G Q G P R R	Q E R S L R G C		large subunit B β
L R Q C Q E Y I K Q Q V S G Q G P R R	Q E R S L R G C		2S ASP of Sharief and Li, 1982
190	200	210	cDNA sequence
C D H L K Q M Q S Q C R C E G L R Q A I E Q Q Q S Q G Q L Q			Ric c 1 large subunit B β
C D H L K Q M Q S Q C R C E G L R Q A I E Q Q Q S Q G Q L Q			data of Machado and Silva, 1992
C D H L K Q M Q S Q C R C E G L R Q A I Q Q Q Q - - - L Q			2S ASP of Sharief and Li, 1982
220	230		cDNA sequence
G Q D V F E A F R T A A N L P S M C G V S P T E C R F			Ric c 1 large subunit B β
G Q D V F E A F R T A A N L P S M C G V S P T E C R F			data of Machado and Silva, 1992
G Q N V F E A F R T A A N L P S M C G V S P T Q C R F			2S ASP of Sharief and Li, 1982

Computer Sequence Analysis

Sequence comparison was performed at the National Centre for Biotechnology Information using the PLAST network service, the SwissProt sequences, EMBL data library, Genbank sequence data bank and Kabat sequences of proteins of immunological interest [19]. The SIM, Local Similarity Program, was used to find the best non-intersecting alignments between two sequences or within a sequence using dynamic programming techniques. The alignments are reported in the order of the decreasing similarity score and share no aligned pairs [20].

Results

Separation of the Subunits A α and A β , B α and B β and End Group Determination

It was shown that Ric c 3 has a native molecular weight of 12 kDa and is composed of two subunits, and Ric c 1 has a native molecular weight of 11 kDa and is composed also of two subunits (see fig. 1). The subunits A α , A β , B α and B β were separated after reduction and S-alkylation using 4-vinylpyridine on a reverse phase C₁₈ column and resolved us-

ing a water/trifluoroacetic acid/acetonitrile gradient (see fig. 2). The NH₂ termini of the A β and B β subunits are blocked as no amino acid was released on Edman degradation, whereas glutamic acid was identified as the NH₂-terminal amino acid for the A α subunit and proline was identified as NH₂-terminal for the B α subunit.

The Primary Structure of the Small Subunits A α and B α

The complete amino acid sequence of the A α chain with 37 amino acid residues is shown in figures 3 and 4. The amino acid sequence data indicate that the small subunit (A α) of the Ric c 3 has glutamic acid as the NH₂-terminal and arginine as the COOH-terminal residue. The molecular mass calculated on the basis of sequence analysis (4,362 daltons) compares well with that determined by SDS-PAGE (5,000 daltons), and the amino acid sequence agreed well within residues 15–51 deduced from the cDNA gene. The complete sequence of the small subunit (B α) of Ric c 1 identified 34 amino acids corresponding to residues 136–169 in the cDNA sequence and a molecular mass of 3,826 daltons. Both sequences were obtained from a single run without

cleavage of the proteins. The peptide sequence starting with proline and ending with arginine was obtained for B α . After 37 and 34 degradation cycles no further amino acids could be identified in A α and B α , respectively.

Determination of the Primary Structure of the Large Subunits A β and B β

The complete amino acid sequences of A β (large subunit of Ric c 3) and B β (large subunit of Ric c 1) consisted of 67 and 65 residues, respectively, and are shown in figures 3 and 4. Since the N-terminal amino acids of A β and B β subunits are blocked and not accessible to direct Edman degradation, subunits A β and B β were sequenced after enzymatic cleavage. The cleavage pattern is given in figure 4. The complete amino acid sequence of A β and B β was derived from tryptic, staphylococcal V8 protease and pyroglutamate aminopeptidase cleavage of the large subunits A β and B β . Digestion of the A β subunit with pyroglutamate aminopeptidase yielded the deblocked peptide QQESQQ-, from which most of the sequence of the subunit could be established. The 67-amino acid residues sequenced correspond completely with the sequence deduced from the cDNA clone in positions 66–132 including the site of blocking. The large subunit B β consists of 65 residues and has a molecular mass of about 7,386 daltons whereas A β has a molecular mass of about 7,670 daltons. These Mr values are in good agreement with those seen in Western blot analysis after SDS electrophoresis. The amino acid sequence of the 2S ASP corresponding to code B has been determined by Sharief and Li [10], and reported to consist of 34 amino acid residues for the small subunit and 61 residues for the large subunit, in contrast to our determination of 34 residues for the small subunit and 65 residues for the large subunit of the 2S ASP code B.

Discussion

The study of inhalant allergens from the structural and immunological point of view has been hindered by the lack of suitable methods for their purification and characterization. In this sense, only in the last years, has significant progress been made in the structural knowledge of the 2S ASP of *R. communis*, since it was first identified about 15 years ago [10].

Our study has shown that the 2S ASP composition of *Ricinus* is more complex than previous reports indicated, and for the first time the two components of the 2S ASP were characterized by complete amino acid sequence analysis. A protein of apparently 12 kDa was previously identified as

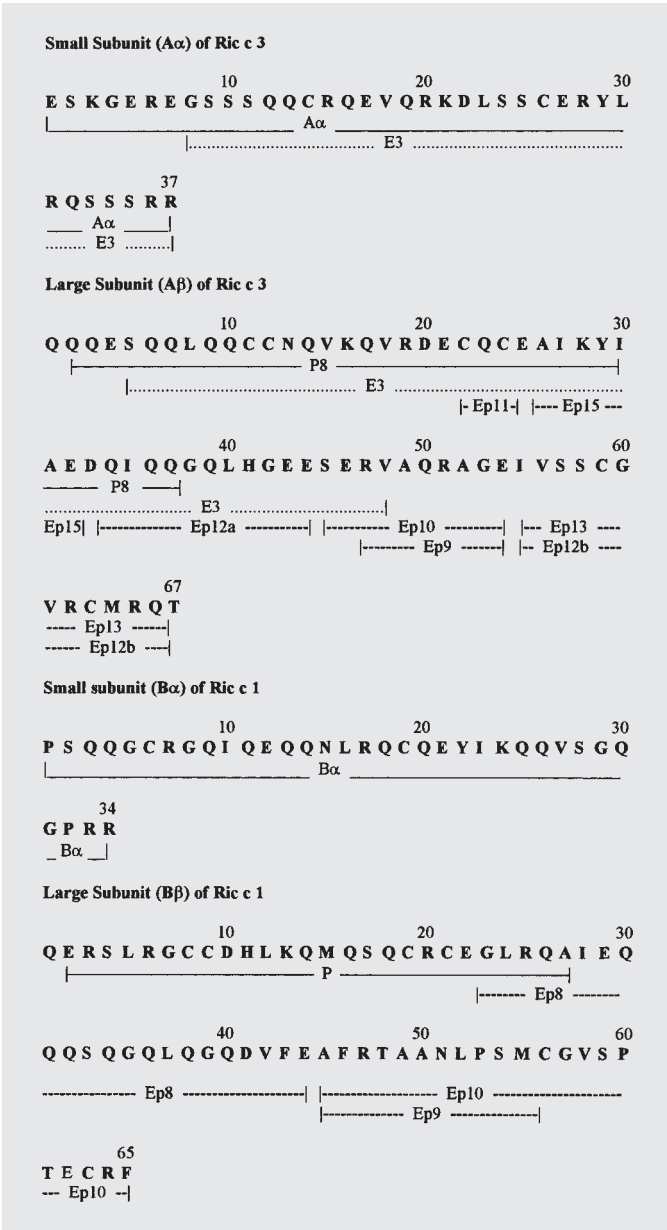


Fig. 4. Cleavage pattern and sequences of Ric c 3 and Ric c 1. The peptides obtained by enzymatic cleavage are marked through lines, as far as they were sequenced. Peptides obtained by *Staphylococcus aureus* V-8, Glu-C endoproteinase are designated E (Ep are only from pyridylethylated subunits). Peptides obtained by cleavage with pyroglutamate aminopeptidase designated P. A α and B α are the small subunits which were sequenced directly without enzymatic cleavage. Ep 12a and Ep12b peptides were sequenced after separation by rechromatography using a modified acetonitrile gradient.

the main protein responsible for asthma outbreaks. This allergen (to which 90% of the patients with asthma attacks are responsive) raised specific IgE antibodies [8, 21, 22].

Ric c 1 consists of two polypeptide chains, a large subunit of 7,386 daltons and a small subunit of 3,836 daltons, linked by disulphide bridges. The comparison of the amino acid sequence of the glutamine-rich storage protein of castor bean (*R. communis*) seeds reported by Sharief and Li [10] as 2S ASP with the Ric c 1 purified in this study reveals a high degree of similarity between the two proteins. The 2S ASP was the entry closest to Ric c 1 when a similarity search based on the amino acid sequence was done. A similarity search in the SwissProt and PIR data banks disclosed 94% similarity between the sequence established for Ric c 1 and the previously reported amino acid sequences for the small and large subunits of 2S ASP [10]. Ric c 1 showed also 100% identity with the region 136–169 and 173–237 in the genome of the precursor protein reported by Irwin et al. [11, 12].

We conclude from the amino acid sequence homology between the 2S ASP of castor beans sequenced by Sharief and Li [10] and Ric c 1 that Ric c 1 purified in this study is identical to the previously sequenced 2S ASP, but we did find some microheterogeneities, too. The amino acid sequence of the 2S ASP, reported by Sharief and Li [10], is similar to our Ric c 1 except for the following differences: (1) difference in amidation states of 2 glutamyl and 1 aspartyl residues at position 29, 41 and 62 of the 2S albumin large subunit, and (2) we have confirmed the sequence -SQGQ- at positions 33–36 in the Ric c 1 (B β) sequence corresponding to positions 205–208 in the cDNA sequence. This tetrapeptide is missing in the 2S ASP sequence of Sharief and Li [10]. In addition, mass spectrometry analysis of B α and B β and pyroglutamate aminopeptidase fragments B β from Ric c 1 of *Ricinus* confirmed the entire sequence and its integrity (unpubl. data). This is in full agreement with the sequence predicted from the cDNA reported by Irwin et al. [11, 12]. The reasons for the differences in the sequences of the Ric c 1 reported here and of the 2S ASP reported by Sharief and Li [10] are unclear. At least in part they seem to be dependent on the variety of castor bean examined (unpubl. data).

To our knowledge, the complete amino acid sequence for Ric c 3 has not been reported. Here we present the first complete amino acid sequence of a Ric c 3 protein, whose existence has been previously suggested from the sequence of the 2S ASP precursor gene [11, 12]. Our sequence of the Ric c 3 subunits corresponds exactly to regions 15–51 and 66–132 as predicted from the cDNA clone of the preproprotein sequence of the 2S albumin gene [11, 12].

Two lines of evidence support this conclusion. The experimental evidence comes from Machado and Silva [13] who isolated a 1.9 S albumin having allergenic activity and denoted it Ric c 3. The protein contained approximately 98 amino acid residues. Ric c 3 was aligned, mainly on the basis of amino acid composition and partial amino acid sequence data, with residues 18–50 and 66–130 of the 2S albumin precursor predicted by the cDNA data [11, 12]. The second piece of evidence for the existence of the heterodimer proteins Ric c 3 and Ric c 1 is presented in this study. Our data differ from those of Machado and Silva [13] with respect to the size of the small subunit which includes residues 15–51 of the cDNA-derived sequence and of the large subunit which composes residues 66–132 of the gene sequence. Thus the complete protein has 104 amino acids.

In describing the allergens Ric c 1 and Ric c 3, we have used the IUIS nomenclature [16] which was revised in 1994 [17]. These allergens were found to be significant components of the *Ricinus* allergens. One of the major *Ricinus* allergens, Ric c 1, was previously considered to be identical to the entire 2S ASP. Because the 2S ASP of castor beans is a mixture of the two proteins 2S ASP code A and 2S ASP code B, Ric c 1 needs to be redefined. We suggest that the nomenclature of the proteins 2S ASP code A and 2S ASP code B should be termed Ric c 3 and Ric c 1, respectively.

The question remains whether the related allergenicity of these proteins can be rationalized in terms of their structural relatedness. Due to their high degree of sequence homology, the complete amino acid sequences obtained confirmed the hypothesis that these allergens and albumin storage proteins in many plant species are closely related compounds. In addition certain structural features are identical, for example the eight cysteine residues which form the four disulphide bridges. Sequence comparison, performed at NCBI using the PLAST network service, the Swiss Prot sequences, EMBL data library, Genbank sequence data bank and Kabat sequences of proteins of immunological interest showed a high degree of homology between Ric c 1, Ric c 3, 2S albumin precursor and other allergens and 2S ASP (tables 1, 2) from other plants. The two heterodimeric 2S ASP of castor bean are obviously members of the family of heterodimeric cysteine-rich proteins found throughout many plant genera.

The sequence homology between the large subunits of Ric c 1 and Ric c 3 is 35%, that of their small subunits is 37%. A comparison with 2S albumins of other genera shows that only the large subunit of Ric c 3 is less homologous to those of the other storage proteins. The small subunit of Ric c 3 and both subunits of Ric c 1 show homology degrees up to 50% with the 2S storage proteins from the

Table 1. Sequence alignment of the small subunits of *R. communis* 2S albumins and allergens from other plants

Small subunits sequences	5	10	15	20	25	30
α of Ric c 3 Allergen	E S K G E R E G S S S Q Q C R Q E V Q - R K D L S S C E R Y					
β of Ric c 1 Allergen		P S Q Q G C R G Q I Q E Q Q N L R Q C Q E Y				
Brazil-Nut 2S albumin †			Q E E C R E Q M Q R Q Q M L S H C R M Y			
Mabinlin IV ‡			E P L C R R Q F Q - Q Q H L R A C Q R Y			
Napin 2 (Rape) §		A G P F R I P K C R K E F Q Q A Q H L R A C Q Q W				
Sin a 1 Allergen ¶		P A G P F R I P K C R K E F Q Q A Q H L R A C Q Q W				
Bra j 1E Allergen **		A G P F R F P R C R K E F Q Q A Q H L R A C Q Q W				
Trypsin inhibitor		P A G P F R L P R C R K E F Q Q A Q H L R A C Q Q W				
Sunflower 2S albumin	E T E I Q R P - - - - E C R F V E Q Q M Q Q P R S - - - -					
			* * \$ * \$ \$ \$ * \$			

Small subunits sequences	35	40
α of Ric c 3 Allergen	L R - Q - S - S - S - - R R	
β of Ric c 1 Allergen	I - K Q Q V - S G Q G P R R	
Brazil-nut 2S albumin	M R - Q Q M E E S	
Mabinlin IV	L R R R A Q R - G	
Napin 2 (Rape)	L H K Q A M Q S G G G	
Sin a 1 Allergen	L H K Q A M Q S G S G P S	
Bra j 1E Allergen	L H K Q A M Q S G S G P	
Trypsin inhibitor	L H K Q A R Q S G S G P S	
Sunflower 2S albumin	- - - - - - - - - - -	
	\$ \$ \$ \$ \$ \$ \$ \$ \$ \$	

The complete sequences of the small subunits of Ric c 3 and Ric c 1 are from this study. Numbering starts from the first amino acid of the small subunit of the mature product Ric c 3. The alignment starts with residue 1 for Ric c 3 and residue 9 for Ric c 1. Gaps indicated by hyphens were added to maximize sequence homology. * = Residues of Ric c 3 and Ric c 1 identical with those of other 2S albumins and allergens; \$ = the fact is indicated that amino acid residues of Sin a 1 and Bra j 1 are identical to the corresponding amino acid residue in Ric c 1 or Ric c 3 (to ease the recognition of conserved amino acid residues, identities are indicated in bold); † = 2S sulphur-rich protein from *Bertholletia excelsa* H. B. K. [24]; ‡ = sweet protein mabinlin IV from *Capparis masaiikai* [29]; § = Napin 2 from rapeseed (*Brassica napus*) [23, 30]; ¶ = major allergen of yellow mustard (*Sinapis alba* L.) [31]; ** = oriental-mustard allergen (*Brassica juncea*) [32].

other genera. The homology between large and small subunits of the same 2S storage proteins varies from 25% (Mabinlin) to 44% (Ric c 3) as is evident from table 3. Interestingly, a closer comparison of small and large subunits revealed several incomplete repeats in each of the examined 2S storage proteins. In Ric c 1, we found three incomplete repeats, namely a copy in region 22–39 and 4–21, a copy in region 14–34 and 9–29 and a copy in region 22–40 and 14–32 in the large subunit and the small subunit, respectively, in addition to that reported by Ericson et al. [23]. Ric c 3 and the other 2S storage proteins revealed between 3 and 4 incomplete repeats each (data not shown).

Thus our data lend additional support to the view that the heterodimeric 2S ASP seem to have evolved from a common genetic origin [24, 25].

All known 2S albumins are synthesized as single precursor proteins that are proteolytically cleaved with the loss

of a linker peptide and short peptides from both the N and C termini [23, 26]. It is noteworthy that the cleavage pattern of Ric c 1 and Ric c 3 which is evident from figure 4 is different from all other 2S albumin processing schemes reported so far. The sequence RR/S terminates both small subunits of Ric c 3 and Ric c 1 resembling signal sequences of prohormone cleavage [27]. The origin of both large subunits through cleavage EN/Q and DN/Q suggests another common protease. Neither these cleavage sites nor the remaining sites ID/E and TN/P occur in the known 2S albumin precursors. Thus, at least three endoproteases seem to be involved in the processing of Ric c 1 and Ric c 3.

Within the group of *R. communis* allergens, Ric c 1 and Ric c 3 fulfil two of the criteria for isoallergens, namely, a similar biological function and molecular size. The third criterion, i.e. 67% identity of amino sequence, is not strictly fulfilled since we find 33% homology. Also, substitutions

Table 2. Sequence alignment of the large subunits of *R. communis* 2S albumins and allergens from other plants

Large subunits sequences	5	10	15	20	25
Aβ of Ric c 3 Allergen	Q Q Q E S	Q Q L Q Q C C N Q	V K Q V R D E C Q C E		
Bβ of Ric c 1 Allergen	Q E R S L R G C C D H L K Q M Q S Q C R C E				
Brazil-Nut 2S albumin †	P R R G M E P H M S E C C E Q L E G M D E S C R C E				
Mabinlin IV ‡	E Q R G P A L R L C C N Q L R Q V N K P C V C P				
Napin 2 (Rape) §	P Q Q R P P L L Q Q C C N E L H Q E E P L C V C P				
Sin a 1 Allergen ¶	P Q Q G P Q Q R P P L L Q Q C C N E L H Q E E P L C V C P				
Bra j 1E Allergen **	Q P Q G P Q Q R P P L L Q Q C C N E L H Q E E P L C V C P				
Trypsin inhibitor	P Q Q G P Q Q R P P L L Q Q C C N E L H Q E E P L C V C P				
Sunflower 2S albumin	Q Q Q R G L Q - Q Q C C N E L Q N V K R E C H C E				
	\$ \$	\$ \$ \$ * * \$	\$ \$		* *

Large subunits sequences	30	35	40	45	50	55
Aβ of Ric c 3 Allergen	A I K Y I A E D Q I Q Q G Q L H G E E S E R V A Q R A G E I					
Bβ of Ric c 1 Allergen	G L R Q A I E Q Q Q S Q G Q L Q G Q D V F E A F R T A A N L					
Brazil-Nut 2S albumin	G L R M M - R M Q Q E E M Q P R G E Q M R R M M R L A E N I					
Mabinlin IV	V L R Q A A H Q Q L Y Q G Q I E G - Q V R R L F R A A R N L					
Napin 2 (Rape)	T L K G A S K A V K Q Q I Q Q Q G Q Q - S R I Y Q T A T H L					
Sin a 1 Allergen	T L K G A S K A V K Q Q V R Q Q L E - I S R I Y Q T A T H L					
Bra j 1E Allergen	T L K G A S K A V K Q Q I R Q Q G - E I S R I Y Q T A T H L					
Trypsin inhibitor	T L K G A A K A V K Q - G Q Q L Q H E I R R I Y Q T A T H L					
Sunflower 2S albumin	A I Q E V A R R V M R Q P Q - Q - Q Q Q Q R R - Q T A - N L					
	\$ \$ \$	\$ \$	\$ \$ \$ \$	\$ -	\$ \$ *	\$

Large subunits sequence	60	65	70	75
Aβ of Ric c 3 Allergen	V S S C G V - - - - R C M R Q T			
Bβ of Ric c 1 Allergen	P S M C G V S P T - E C R F			
Brazil-Nut 2S albumin	P S R C N L S P M - R C P M G G S			
Mabinlin IV	P N I C K I P A V G R C Q F T R W			
Napin 2 (Rape)	P K V C N I P Q V S V C P F Q K T M P G			
Sin a 1 Allergen	P K V C N I P Q V S V C P F K K T M P G P S			
Bra j 1E Allergen	P R V C N I P R V S I C P F Q K T M P G P S			
Trypsin inhibitor	P K V C N I P Q V Q V C P F N K T M P G P S			
Sunflower 2S albumin	P N Q C D L E - V Q Q C T T C T G			
	\$ *	* \$ \$		

The complete sequences of the large subunits of Ric c 3 and Ric c 1 are from this study. Numbering starts from the first amino acid of the large subunit of the mature product Ric c 3. The alignment starts with residue 1 for Ric c 3 and residue 4 for Ric c 1. Gaps indicated by hyphens were added to maximize sequence homology. * = Residues of Ric c 3 and Ric c 1 identical with those of other 2S albumins and allergens; \$ = the fact is indicated that amino acid residues of Sin a 1 and Bra j 1 are identical to the corresponding amino acid residue in Ric c 1 or Ric c 3 (to ease the recognition of conserved amino acid residues, identities are indicated in bold); † = 2S sulphur-rich protein from *Bertholletia excelsa* H. B. K. [24]; ‡ = sweet protein mabinlin IV from *Capparis masaiikai* [29]; § = Napin 2 from rapeseed (*Brassica napus*) [23, 30]; ¶ = major allergen of yellow mustard (*Sinapis alba* L.) [31]; ** = oriental-mustard allergen (*Brassica juncea*) [32].

of basic residues by basic ones, or acidic by acidic ones are not frequent, altogether two. However, if one considers all polar residues as similar, the homology is raised to 62%. In our opinion this may be regarded as sufficient criterion to place the two proteins Ric c 1 and Ric c 3 into the group of isoallergens, since most frequently, antigenic determinants

are rich in polar residues situated in accessible regions of antigenic protein molecules.

As discussed above, considerable homology is found between the members of individual families of albumin storage proteins. The members of each are related and they may have arisen from a common ancestral gene. The present

Table 3. Homology relationships between Ric c 3, Ric c 1 and 2S albumins from other genera

Family	Amino Acid Homology %													
	<i>Euphorbiaceae</i>				<i>Lecythidaceae</i>		not yet classified		<i>Brassicaceae</i>					
	<i>Ricinus communis</i>				<i>Bertholletia excelsa</i> H.B.K.		<i>Capparis masaiikai</i> Levl.		<i>Brassica napus</i>	<i>Sinapis alba</i> L.	<i>Brassica juncea</i>	<i>Sinapis arvensis</i>		
	<i>Ric c 3</i> Allergen		<i>Ric c 1</i> Allergen		Brazil-Nut 2S Albumin		Mablinin IV (Sweet Protein)		Napin 2 (Rape)	<i>Sin a 1</i> Allergen	<i>Bra j 1E</i> Allergen	Trypsin inhibitor (Charlock)		
	S	L	S	L	S	L	S	L	S	L	S	L	S	L
<i>Ric c 3</i> Allergen	44% (7/16)		37% (10/27)	35% (20/58)	46% (10/22)	29% (17/58)	44% (10/23)	33% (18/55)	30% (7/23)	33% (20/60)	30% (7/23)	33% (18/55)	30% (7/23)	37% (22/60)
<i>Ric c 1</i> Allergen			41% (7/17)		50% (12/24)	40% (25/62)	50% (11/22)	47% (26/55)	39% (10/26)	37% (19/54)	39% (10/38)	32% (16/54)	39% (10/26)	32% (17/54)
Brazil-Nut 2S Albumin					43% (12/28)		44% (10/23)	36% (20/56)	31% (8/26)	27% (18/56)	31% (8/26)	23% (13/56)	32% (8/25)	25% (14/56)
Mablinin IV (Sweet Protein)							25% (6/24)		58% (15/26)	43% (26/70)	58% (15/26)	37% (26/70)	58% (15/26)	43% (30/70)
Napin 2 (Rape)									30% (6/20)		97% (35/36)	91% (68/75)	92% (33/36)	87% (65/75)
<i>Sin a 1</i> Allergen											30% (6/20)		95% (35/37)	88% (70/80)
<i>Bra j 1E</i> Allergen													33% (5/15)	86% (69/80)
Trypsin inhibitor (Charlock)														35% (7/20)

Percent homology was calculated from the number of similar residues per length of residues overlapped (in parentheses), with gap frequency 0.0%, using the sequence analysis programme SIM – Local Similarity Program [20]. Percent homology of small subunits (S) and percent homology of large subunits between different proteins compared (L) are shown. Numbers in bold show homology between large and small subunits of the same protein.

finding of sequence homology between storage proteins and a trypsin inhibitor may not only suggest an interesting divergent evolution of two different classes of seed proteins but also support a view that proteinase inhibitors also act as storage proteins. This homology may lead to the question as to whether there is any functional relationship between allergens, enzyme inhibitors and plant storage proteins.

Despite possible differences in their subunit structure and biosynthesis, all the 2S albumins seem to be compact globular proteins with conserved cysteine residues. Although little is known about the detailed three-dimensional structure of 2S albumins, that of yellow mustard has been reported to contain ~50% α -helix, with little or no β -sheet [28]. Therefore, the possibility exists that some common structural features might be associated with their capacity to induce a hypersensitivity reaction.

Experiments are currently in progress to address the relative abundance of each family member in castor bean, as mentioned above. It is equally important to determine the structural basis of the immunogenicity and allergenicity of the 2S albumin family members.

Acknowledgements

This work was supported by a grant from the Deutsche Akademische Austauschdienst to M. E. H. Bashir. We thank Dr. Renate Berger-Hoffmann and Dr. Ralf Hoffmann for help in the initial phase of this work, Norbert Weber and Wolfgang Reinle for technical support, Dr. Kerstin Strupat, University of Münster, for mass-spectrometric analyses and Professor Michael Ashwort for linguistic corrections.

References

- 1 Layton LL, Dante BT, Moss LK, Dye NH, DeEds F: Electrophoretic fractionation of soluble antigenic proteins from the seed of *Ricinus communis* (castor bean). J Am Oil Chem Soc 1961;38:405–410.
- 2 Panzani RC, Layton LL: Allergy to the dust of *Ricinus communis* (castor bean): Clinical studies upon human beings and passively sensitized monkeys. Int Arch Allergy Appl Immunol 1963;22:350–368.
- 3 Layton LL, Panzani RC, Von Helms L, Campell CK: Cross-reactivity in primary respiratory allergy to castor bean (*Ricinus communis*). Int Arch Allergy 1970;37:67–75.

- 4 Spies JR: Allergens. J Agric Food Chem 1976; 22:30–36.
- 5 Thorpe SC, Kemeny DM, Panzani RC, McGurl B, Lord M: Allergy to castor bean. II. Identification of the major allergens in castor bean seeds. J Allergy Clin Immunol 1988;82 1:67–72.
- 6 Kemeny DM, Frankland AW, Fahkri ZI, Trull AK: Allergy to castor bean in the Sudan: Measurement of serum IgE and specific IgE antibodies. Clin Allergy 1981;11:463–471.
- 7 Lehrer SB, Karr RM, Salvaggio JR: Extraction and analysis of coffee bean allergens. Clin Allergy 1978;8:217–226.
- 8 Lehrer SB, Karr RM, Salvaggio JR: Analysis of green coffee bean and castor bean allergens using RAST inhibition. Clin Allergy 1981;11: 357–366.
- 9 Youle RJ, Huang AHC: Evidence that the castor bean allergens are the storage proteins in the proteins bodies of castor bean. Plant Physiol 1978;61:1040–1042.
- 10 Sharief FS, Li SS: Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. J Biol Chem 1982; 257:14753–14759.
- 11 Irwin SD, Lord JM: Nucleotide sequence of a *Ricinus communis* 2S albumin precursor gene. Nucleic Acids Res 1990;18:5890.
- 12 Irwin SD, Keen JN, Findlay JBC, Lord JM: The *Ricinus communis* 2S albumin precursor: A single preproprotein may be processed into two different heterodimeric storage proteins. Mol Gen Genet 1990;222:400–408.
- 13 Machado OLT, Silva JG Jr: An allergenic 2S storage protein from *Ricinus communis* seeds which is a part of the 2S albumin precursor predicted by cDNA data. Braz J Med Biol Res 1992;25:567–582.
- 14 Spies JR, Coulson EJ: The chemistry of allergens. VIII. Isolation and properties of an active protein-polysaccharidic fraction, CB-1A, from castor beans. J Am Chem Soc 1943;65:1720–1725.
- 15 Bashir MEH: Immunological and Structural Studies of Ric c 1 and Ric c 3, the Major Allergens of *Ricinus communis*, var. Pacific-6; PhD thesis; University of Saarland, 1997.
- 16 Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TAE: Allergen nomenclature. J Allergy Clin Immunol 1987;80:639–645.
- 17 King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TAE, Thomas W: Allergen nomenclature. Bull World Health Organ 1994;72: 797–806.
- 18 Aitken A, Geisow MJ, Findlay JB, Holmes C, Yarwood A: Peptide preparation and characterisation: Protein sequencing; a practical approach; in Rickwood D, Hames BD (eds): Practical Approach Series. Oxford, IRL Press at Oxford University Press, 1988, chap 2, pp 43–68.
- 19 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990;215:403–410.
- 20 Huang X, Miller W: A time-efficient, linear-space local similarity algorithm. Adv Appl Math 1991;12:337–357.
- 21 Panzani RC: La storia dell'allergia al ricino (*Ricinus communis*), modello tipico di sensibilizzazione reaginica umana. Asma Allergia Immunopatol 1990;79:5–12.
- 22 Thorpe SC, Kemeny DM, Panzani RC, Lessof MH: The relationship between total IgE and castor bean-specific IgE antibodies in castor bean-sensitive patients from Marseilles. Int Arch Allergy Appl Immunol 1987;82:456–460.
- 23 Ericson ML, Rödin J, Lenman M, Glimelius K, Josefsson L-G, Rask L: Structure of the rape-seed 1.7 storage protein, Napin, and its precursor. J Biol Chem 1986;261:14576–14581.
- 24 Ampe C, Van Damme J, de Castro LAB, Sampaio MJAM, Van Montagu M, Vandekerckhove J: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). Eur J Biochem 1986;159: 597–604.
- 25 Odani S, Koide T, Ono T, Ohnishi K: Structural relationship between barley (*Hordeum vulgare*) trypsin inhibitor and castor-bean (*Ricinus communis*) storage protein. Biochem J 1983; 213:543–545.
- 26 Crouch ML, Tenbarge KM, Simon AE, Ferl R: cDNA clones for *Brassica napus* seed storage proteins: Evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J Mol Appl Genet 1983;2:273–283.
- 27 Jörnvall H, Carlström A, Pettersson T, Jacobsson B, Persson M, Mutt V: Structural homologies between prealbumin, gastrointestinal prohormones and other proteins. Nature 1981;291: 261–263.
- 28 Menendez-Arias L, Monsaive RI, Gavilanes JG, Rodriguez R: Molecular and spectroscopic characterisation of a low molecular weight seed storage protein from yellow mustard (*Sinapis alba* L.). Int J Biochem 1987;19:899–907.
- 29 Nirasawa S, Nishino T, Katahira M, Uesugi S, Hu Z, Kurihara Y: Structure of heat-stable and unstable homologues of the sweet protein mabinlin. The difference in the heat stability is due to replacement of a single amino acid residue. Eur J Biochem 1994;223:989–995.
- 30 Josefsson L-G, Lenman M, Ericson ML, Rask L: Structure of a gene encoding the 1.7 S storage protein, Napin, from *Brassica napus*. J Biol Chem 1987;262:12196–12201.
- 31 Menendez-Arias L, Moneo I, Dominguez J, Rodriguez R: Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, *Seda* I. Eur J Biochem 1988;177:159–166.
- 32 Monsalve RI, Gonzalez de la Pena MA, Menendez-Arias L, Lopez-Otin C, Villalba M, Rodriguez R: Characterisation of a new oriental-mustard (*Brassica juncea*) allergen, *Bra j* IE: Detection of an allergenic epitope. Biochem J 1993; 293:625–632.

Copyright: S. Karger AG, Basel 1998. Reproduced with the permission of S. Karger AG, Basel. Further reproduction or distribution (electronic or otherwise) is prohibited without permission from the copyright holder.