

Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, *Sin a* I

Luis MENÉNDEZ-ARIAS¹, Ignacio MONEO², Javier DOMÍNGUEZ² and Rosalía RODRÍGUEZ¹

¹ Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Complutense, Madrid

² Servicio de Alergia e Inmunología, Centro Ramón y Cajal, Madrid

(Received April 11/June 21, 1988) — EJB 88 0409

Sin a I, a 2-S albumin from the seeds of yellow mustard, is herein described as the major allergen of these seeds. This protein is composed of two disulfide-linked polypeptide chains of 39 and 88 amino acids, whose primary structures are reported. The *Sin a* I allergen is found to be related to other low-molecular-mass albumins, such as those isolated from rapeseed, castor bean and Brazil nut. Additional structural similarity has also been found between the glutamine-rich large chain of *Sin a* I and a proline-rich zein, a gliadin, and trypsin and α -amylase inhibitors isolated from the seeds of several monocotyledons. Internal amino acid sequence similarity has been detected at both termini of the small and large chains of *Sin a* I and involves the location of proline and glycine residues at similar positions in relation to the processing cleavage sites. Prediction of secondary structure, based on the amino acid sequences of the mature chains of the mustard allergen, indicates that the precursor polypeptide is cleaved at regions showing a high β -turn probability. This is also observed with the amino acid sequence deduced from the rapeseed napin gene nucleotide sequence.

Seed storage proteins constitute more than 10% of the total protein isolated from seeds and accumulate in aggregates called protein bodies. Their main function is the provision of an adequate nitrogen supply to the germinating seedling and they do not usually show any enzymatic activity. Many investigations on these kind of proteins have been carried out during the last two decades, mainly on their chemical composition and nutritional value. Nevertheless, seed storage proteins are also of interest in clinical research. Thus, gliadins, or peptides derived from them, initiate damage to the absorptive epithelium of the small intestine during ingestion, producing symptoms of celiac disease in susceptible individuals [1]. Allergenic proteins have also been found in a wide variety of seeds including castor bean, soybean, cottonseed, pea, etc. [2]. These allergens are frequently low-molecular-mass proteins with a high cystine content, but information on their biochemical and structural properties is still scarce. In some cases, allergenicity has been ascribed to the 1.7–2.0-S albumin fraction, as in cottonseed or soybean [3, 4], but other seed allergens fail to meet this rule (e.g. peanut allergens) [5].

An increasing interest in Brassicaceae and particularly in rapeseed (*Brassica napus* L.) proteins as a consequence of their being one of the most important annual oilseed crops in the world, has recently allowed their characterization and elucidation of their gene nucleotide sequences [6–9]. Rapeseed proteins are composed of two major types. One is a 12-S neutral globulin, apparently similar to the 11–12-S 'legumin-like' storage proteins found in seeds of many other

species [7, 10]. The other major proteins are small basic albumins, referred to as the 1.7-S storage proteins [11] and which have also been called napins [6, 12]. However, rapeseeds are not used as human food and no allergy related to these proteins has been reported. Among those cruciferous plants which are suspected to be allergenic, mustards are the most relevant, though cases of cabbage allergy have also been reported [13]. Table mustard is usually a blend of flour obtained from the seeds of two different species: yellow (*Sinapis alba* L.) and oriental (*Brassica juncea* L. Czern & Coss.) mustard. The relative amounts of both types of flour varies according to the manufacturer, but the yellow is the most frequently found. In a previous paper, we reported the molecular and spectroscopic characterization of the 1.7-S storage protein of the yellow mustard seeds [14]. We have observed that this fraction is the most allergenic for atopic individuals. Thus, the aim of this paper was to elucidate the primary structure of the major allergen of *Sinapis alba* L. seeds as a basis for further investigations on the structure allergenicity relationship of this protein.

EXPERIMENTAL PROCEDURES

Portions of this article, including Experimental Procedures, Figs 1 and 2 and Tables 1–6, are presented in a miniprint section at the end of this paper.

RESULTS

Allergenic properties

Clinical screening performed with sera of individuals suspected to be sensitive to food allergens revealed that about 16.8% of them had specific IgE antibodies against mustard allergens. These data are in agreement with previously reported evidence on the incidence of spice allergy in other countries

Correspondence to R. Rodríguez, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Complutense, Ciudad Universitaria, E-28040 Madrid, Spain

Enzymes. Trypsin (EC 3.4.21.4); chymotrypsin (EC 3.4.21.1); thermolysin (EC 3.4.24.4); carboxypeptidase Y (EC 3.4.16.1); L-pyroglutamyl-peptide hydrolase (EC 3.4.19.3); clostridiopeptidase B (EC 3.4.22.8).

[21, 22]. However, no further structural studies have been performed on these seed allergens. If the allergenic extract prepared according to the procedure described in the Supplement, is applied to a Sephadex G-75 column, three main fractions can be collected, which correspond to the 12-S globulins, the 1.7-S albumins and small molecules eluting at the total volume. The elution profile and the allergenicity of the different fractions (measured by inhibition reverse-enzyme immunoassay) are given in Fig. 1 (Supplement). The 1.7-S fraction seemed to be more allergenic than the 12-S globulins, according to these data. Furthermore, leukocyte histamine-release assays, performed with the sera of eight patients showing atopic allergy to mustard, revealed that the 1.7-S albumin fraction was about 3–4 times more active than the 12-S globulins. Radio-allergosorbent inhibition tests indicated that more than 60% of the IgE in sera from atopic individuals was inhibited by 10 µg of the 1.7-S seed albumin, when assayed on mustard crude extract, previously coupled to BrCN-activated paper discs. On the other hand, the same amount of protein from the 12-S globulin fraction was unable to produce an inhibition greater than 10%, indicating that most of the allergenic activity of this food resides in to the low-molecular-mass albumins (data not shown).

From the results given above, it could be concluded that the 1.7-S albumin fraction of the yellow mustard seeds is the major allergen in this biological material. It may be termed as *Sin a I*, according to the proposals of the IUIS Sub-Committee for Allergen Nomenclature, meeting the purity requirements necessary in order to be designated according to the new system [23]. Moreover, this protein shows some biochemical properties also found in other allergens, such as, low molecular mass, resistance to proteolytic treatments and a high cystine content.

Amino acid sequence of the small chain

Chymotryptic digestion of the reduced and carboxyamidomethylated small chain was fractionated by reversed-phase chromatography (Fig. 2a). The best profiles were obtained after a 3-h hydrolysis at an enzyme/substrate ratio of 1:500 (by mass). Under these conditions, peptide bonds involving the carboxyl group of methionine were not cleaved. If the amount of chymotrypsin is increased to a ratio of 1:100 (by mass), the peptide CT-5 is cleaved at the Met–Gln bond and other peptides are also generated by cleavage of trypsin-susceptible bonds. The amino acid compositions of the chymotryptic peptides CT-1 to CT-5 are shown in Table 1. All were recovered at a similar yield. The recoveries of CT-2a and CT-2b were summed since these two peptides are nearly identical. In fact, their amino acid sequence showed that the only difference between them is at the NH₂-terminal residue which is arginine in CT-2a and glycine in CT-2b. These data suggest the presence of at least two clones for this protein. Other heterogeneities involving deamination of glutamine residues are suspected but these may be a consequence of the functional role of the protein and are probably not an original mutation.

All the chymotryptic peptides were manually sequenced except CT-1 which was extracted in the organic phase of the first cycle of the Edman degradation due to its highly hydrophobic character. The correct positioning of amino acids was established by comparison with both the NH₂-terminal sequence of the whole polypeptide chain and the analysis of the structure of T-1.

Total or partial amino acid sequences and compositions of several tryptic peptides (T-1, T-2, T-3, T-5 and T-6) were useful in order to obtain the correct alignment of the chymotryptic peptides (Fig. 3). The elution profile and the amino acid compositions of the resulting tryptic peptides are shown in Fig. 2b and Table 2, respectively.

The COOH-terminal of the sequence was not fully resolved by analysis of the tryptic and chymotryptic digests and, thus, a cyanogen-bromide treatment was used to complete the sequence. After a 20-h hydrolysis, a peptide containing three serines, two glycines, one glutamine and one proline was obtained. However, direct sequencing was not possible, as glutamine undergoes cyclisation under acidic conditions and most of this peptide was blocked at the amino-terminal. Digestion with 5-oxopropyl peptidase (L-pyrroglutamylpeptide hydrolase) was used to remove the pyrroglutamyl residue and the resulting hexapeptide (CN-1-SOP) was sequenced. The primary structure obtained was in agreement with the results of the carboxypeptidase digestion, which released serine after a 24-h treatment.

Amino acid sequence of the large chain

The NH₂-terminal amino acid sequence obtained by automatic Edman degradation identified only 19 residues, due to the high proline content. More than 20% of the residues found in this polypeptide chain are glutamine, which is susceptible cyclisation to pyrroglutamic acid when located at the NH₂-terminal position of peptides, making it difficult to determine by Edman degradation. Thus, thermolytic digestion was the most promising, since no cleavage would be expected at the amino groups of the glutamine residues involved in peptide bonds. The elution profile of the thermolytic digestion is shown in Fig. 2c. Some of the peaks observed represented two or three different peptides and a further step of purification was required. The amino acid composition of the resulting peptides is given in Table 3. Yields ranged from around 40% for the larger peptides up to 75% for dipeptides and tripeptides.

Alignment was completed after tryptic digestion of the reduced and radioalkylated large chain. The elution profile and the amino acid compositions of the resulting peptides are shown in Fig. 2d and Table 4, respectively. Peptide T-5 contains a high number of glutamine residues and thus deamidation processes could become a major problem. In fact, several polypeptides showing the same amino acid composition can be identified (Fig. 2d), revealing that those glutamine residues exhibit a certain tendency towards deamidation. Peptide T-5 was sequenced automatically.

An additional chymotryptic digestion was performed in order to confirm the whole amino acid sequence of the large chain. This proteolytic treatment gave rise to complex reversed-phase chromatograms, due to the high number of cleavage sites in this chain. However, CT-4 was identified by its absorbance at 274 nm; its amino acid sequence allows an overlap at positions 58–59. A chymotryptic peptide including Lys68–Val69, adequate for manual or automatic sequencing, could not be obtained. In addition, this peptide bound was also sensitive to digestion by clostridiopeptidase B. Despite these problems, the proposed alignment is supported by the strong similarity with the sequence of rapeseed napin (Fig. 5).

The amino acid sequences of both chains (large and small) of the *Sin a I* allergen are shown in Fig. 3. It can also be observed that the proposed structure is in good agreement

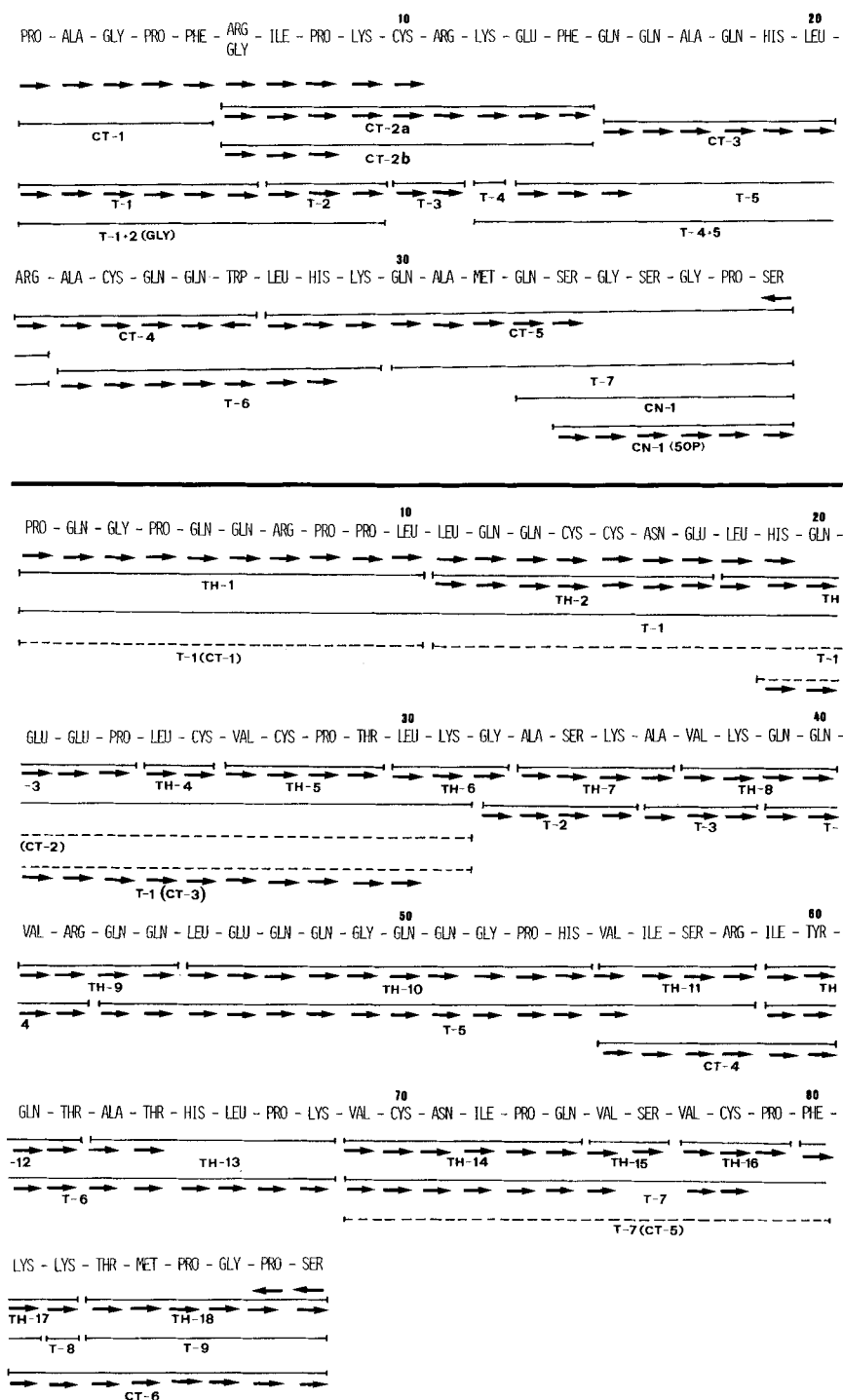


Fig. 3. Complete amino acid sequences of the small (upper) and large (lower) chains of the major allergen of yellow mustard seeds, Sin a I. Peptides obtained by cleavage with trypsin (T), chymotrypsin (CT), thermolysin (TH), cyanogen bromide (CN) or L-pyroglutamyl-peptide hydrolase (5 OP) are indicated by a continuous line. Discontinuous lines represent peptides obtained as subproducts of the enzymatic digestions. Right arrows (→) correspond to residues identified by manual or automatic Edman degradation, while left arrows (←) indicate amino acids determined after digestion with carboxypeptidase Y.

with the amino acid compositions of the native protein and its polypeptide chains reported previously (Table 5).

DISCUSSION

The yellow mustard seed allergen, *Sin a* I, is composed of two polypeptide chains of 39 and 88 amino acids, respectively.

In contrast to other seed proteins [25, 26], its primary structure does not show any tandem repeat of conserved sequences. However, a high glutamine content is observed at positions 39–51 of the large subunit, which could play a role in providing nitrogen for the germinating seedling, in a similar way to that found in other seed storage proteins [27]. In addition, if the primary structures of both *Sin A* 1 chains are aligned, significant structural similarity is found at the NH₂- and

Unfortunately, evidence is still poor, but this could be an intriguing hypothesis to provide light in studies concerning structure/allergenicity relationships.

This work was supported by grant PB86-0530 from the *Dirección General de Investigación Científica y Técnica* (M.E.C., Spain). We want to express our gratitude to Prof. A. Schnitzer (North Dakota State University, USA) who provided us with pure yellow mustard seeds flour, Dr E. Méndez (Centro Ramón y Cajal, Madrid) for technical facilities in the automatic sequence determination, and to Drs J. M. Claverie, L. Bricault and I. Sauvaget (Institut Pasteur, Paris) who kindly sent us their Protein Data Bank. Thanks are also due to Dr J. G. Gavilanes for discussion.

REFERENCES

- Weiss, J. B., Austin, R. K., Schanfield, M. S. & Kagnoff, M. F. (1983) *J. Clin. Invest.* 72, 96–101.
- Spies, J. R. (1974) *J. Agric. Food Chem.* 22, 30–36.
- Youle, R. J. & Huang, A. H. C. (1979) *J. Agric. Food Chem.* 27, 500–503.
- Lemanske, R. F. Jr & Taylor, S. L. (1987) *Clin. Rev. Allergy* 5, 23–36.
- Barnett, D. & Howden, M. E. H. (1986) *Biochim. Biophys. Acta* 882, 97–105.
- Crouch, M. L., Tenbarger, K. M., Simon, A. E. & Ferl, R. (1983) *J. Mol. Appl. Genet.* 2, 273–283.
- Simon, A. E., Tenbarger, K. M., Scofield, S. R., Finkelstein, R. R. & Crouch, M. L. (1985) *Plant Mol. Biol.* 5, 191–201.
- Josefsson, L.-G., Lenman, M., Ericson, M. L. & Rask, L. (1987) *J. Biol. Chem.* 262, 12196–12201.
- Scofield, S. R. & Crouch, M. L. (1987) *J. Biol. Chem.* 262, 12202–12208.
- Derbyshire, E., Wright, D. J. & Boulter, D. (1976) *Phytochemistry* 15, 3–24.
- Lönnnerdal, B. & Janson, J.-C. (1972) *Biochim. Biophys. Acta* 278, 175–183.
- Ericson, M. L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. & Rask, L. (1986) *J. Biol. Chem.* 261, 14576–14581.
- Blaiss, M. S., McCants, M. L. & Lehrer, S. B. (1987) *Ann. Allergy* 58, 248–250.
- Menéndez-Arias, L., Monsalve, R. I., Gavilanes, J. G. & Rodríguez, R. (1987) *Int. J. Biochem.* 19, 899–907.
- Domínguez, J., Cuevas, M., Moneo, I., Ureña, V., Ferreira, A. & Bootello, A. (1984) *Allergy (Copenh.)* 39, 529–533.
- Carrillo, T., Cuevas, M., Muñoz, T., Hinojosa, M. & Moneo, I. (1986) *Contact Dermatitis* 15, 69–72.
- Gavilanes, J. G., González de Buitrago, G., Pérez-Castells, R. & Rodríguez, R. (1982) *J. Biol. Chem.* 257, 10267–10270.
- Tarr, G. E. (1977) *Methods Enzymol.* 47, 335–357.
- Fohlman, J., Rask, L. & Peterson, P. A. (1980) *Anal. Biochem.* 106, 22–26.
- Méndez, E. & Lai, C. Y. (1975) *Anal. Biochem.* 68, 47–53.
- Niinimäki, A. & Hannuksela, M. (1981) *Allergy (Copenh.)* 36, 487–493.
- Thiel, C., Fuchs, E., Maasch, H.-J. & Wahl, R. (1986) in *New trends in allergy II* (Ring, J. & Burg, G. eds) pp. 154–164, Springer-Verlag, Berlin, Heidelberg.
- Marsh, D. G., Goodfriend, L., King, T. P., Löwenstein, H. & Platts-Mills, T. A. E. (1986) *Bull. W. H. O.* 64, 767–770.
- Beaven, G. H. & Holiday, E. R. (1952) *Adv. Protein Chem.* 7, 319–382.
- Pedersen, K., Argos, P., Naravana, S. V. L. & Larkins, B. A. (1986) *J. Biol. Chem.* 261, 6279–6284.
- Kasarda, D. D., Okita, T. W., Bernardin, J. E., Baecker, P. A., Nimmo, C. C., Lew, E. J.-L., Dietler, M. D. & Greene, F. C. (1984) *Proc. Natl Acad. Sci. USA* 81, 4712–4716.
- Argos, P., Pedersen, K., Marks, M. D. & Larkins, B. A. (1982) *J. Biol. Chem.* 257, 9984–9990.
- Lipman, D. J. & Pearson, W. R. (1985) *Science (Wash. DC)* 227, 1435–1441.
- Claverie, J. M. & Bricault, L. (1986) *Proteins Struct. Funct. Genet.* 1, 60–65.
- Ampe, C., van Damne, J., de Castro, L. A. B., Sampaio, M. J. A. M., van Montagu, M. & Vandekerckhove, J. (1986) *Eur. J. Biochem.* 159, 597–604.
- Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* 47, 251–276.
- von Heijne, G. (1983) *Eur. J. Biochem.* 133, 17–21.
- Schwartz, T. W. (1986) *FEBS Lett.* 200, 1–10.
- Mole, L. E., Goodfriend, L., Lapkoff, C. B., Kehoe, J. M. & Capra, J. D. (1975) *Biochemistry* 14, 1216–1220.
- Shipolini, R. A., Callewaert, G. L., Cottrell, R. C. & Vernon, C. A. (1974) *Eur. J. Biochem.* 48, 465–476.
- Elsayed, S. & Bennich, H. (1975) *Scand. J. Immunol.* 4, 203–208.
- Goodfriend, L., Choudhury, A. M., Kappler, D. G., Coulter, K. M., Dorval, G., Del Carpio, J. & Osterland, C. K. (1985) *Mol. Immunol.* 22, 899–906.
- Drenth, J., Low, B. W., Richardson, J. S. & Wright, C. S. (1980) *J. Biol. Chem.* 255, 2652–2655.

Supplementary Material to

Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seeds, *Sin a I*

Luis MENÉNDEZ-ARIAS, Ignacio MONEO, Javier DOMÍNGUEZ and Rosalía RODRÍGUEZ

EXPERIMENTAL PROCEDURES

Determination of allergenic activity

Serum samples were obtained from 18 non-previously hypo-sensitized individuals with suspected food allergy to table mustard. The crude mustard extract was obtained through dilution of delipidized mustard flour in 0.15 M borate buffer pH 8.0 (at 10%, mass/vol.). For further determinations, these extracts were centrifuged at 2000 × g and filtered. Direct and inhibition reverse enzyme immunoassays and leukocyte histamine-release assays were performed as previously described [15, 16]. Histamine release was measured by fluorometry, after reaction with *o*-phthalaldehyde, including internal and external histamine · HCl standards. Radio-allergosorbent tests were performed according to the instructions of the manufacturer (Pharmacia, Uppsala, Sweden). Radio-allergosorbent inhibition assays were carried out by incubating an 0.05-ml aliquot of five

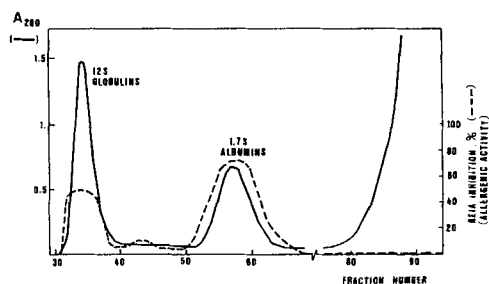


Fig. 1. Gel filtration chromatography of the mustard crude extract (20 mg at 4 mg/ml) on a Sephadex G-75 column (2.6 × 100 cm), equilibrated in 0.15 M ammonium bicarbonate, pH 8.0. The flow rate was 25 ml/h and 4-ml fractions were collected. Inhibition reverse-enzyme immunoassays were performed using 5-μl aliquots, after 1:10 dilution with distilled water. Horseradish peroxidase (0.5 μg) was used in each determination for obtaining the allergenic conjugate

pooled positive sera for 1 h at room temperature. Then, a CNBr-activated paper disc, onto which crude mustard extract was previously coupled, was added to the samples and the radio-allergosorbent test was performed as usual.

Protein purification

Yellow mustard seeds allergen, *Sin a I*, was purified as previously reported [14], and the allergenic activity was measured at the different stages of the process. Separation of the large and small chains of this allergen was achieved by gel filtration chromatography after reduction of disulfide bonds and radioalkylation with iodo- ^{14}C acetamide [14].

Enzymatic and chemical cleavage

Tryptic and chymotryptic digestions of both reduced and carboxyamidomethylated polypeptide chains of *Sin a I* were performed as previously described [17]. Hydrolysis with thermolysin of the reduced and carboxyamidomethylated large chain was carried out at a protein concentration of 4–5 mg/ml in 0.1 M ammonium bicarbonate pH 8.0, containing 1 mM calcium chloride. An enzyme/substrate ratio of 1:100 was used, and the mixture was incubated at 37°C for 2 h. Both carboxyamidomethylated polypeptide chains were also cleaved with carboxypeptidase Y (Boehringer, FRG). For this, 90–100 nmol of each polypeptide was separately dissolved in 0.7 ml 0.1 M pyridine/acetic acid buffer, pH 5.5, containing 10 mM EDTA and an internal standard of norleucine (25 nmol/ml). An enzyme/substrate ratio of 1:50 was used and the mixture was incubated at 37°C. Aliquots were collected at intervals up to 24 h and amino acid analysis was performed to identify the amino acids released.

Cyanogen bromide cleavage was performed by dissolving each reduced and carboxyamidomethylated polypeptide in 70% formic acid, at a concentration of 10 mg/ml. Cyanogen bromide was added to a 40-molar excess over methionine. After flushing with nitrogen, the reaction was allowed to proceed for 20 h at room temperature in the dark. Finally, the sample was diluted with distilled water and lyophilized.

To cleave the amino-terminal 5-oxoproline (pyroglutamic acid) of the CNBr peptide from the small chain, 50 nmol cyclised peptide was dissolved in 0.2 ml 0.1 M sodium phosphate buffer, pH 8.0, containing 10 mM EDTA, 5% (by vol.) glycerol and 5 mM dithiothreitol. Pyroglutamyl-peptide hydrolase (Boehringer, FRG) (0.1 ml of a 10-mg/ml solution) was added, and the reaction mixture maintained at 37°C for 2 h.

Peptide separation

Every digest was lyophilized, dissolved in 0.1% (by vol.) trifluoroacetic acid and centrifuged (in a Beckman microfuge) prior to injection in a HPLC column. 30–100 nmol of digested protein were chromatographed. The separation was performed on a Beckman HPLC system equipped with a Nucleosil 10 C_{18} column (250 mm $\times$ 8 mm $\times$ 4 mm). A gradient of acetonitrile in 0.1% (by vol.) trifluoroacetic acid was employed for elution of peptides. The eluent was continuously monitored at both 214 nm and 280 nm. Radioactive peptides were detected by counting aliquots of 5–20% of the total volume of each collected fraction. Many peptides required a further purification step which was usually achieved through reversed-phase chromatography using a different column and elution system: an Ultrasphere-octadecylsilane (5 μm) column (150 mm $\times$ 4.5 mm) and a gradient of acetonitrile in 10 mM ammonium acetate buffer, pH 6.5

Amino acid analysis

Peptide hydrolysis was carried out for 20 h at 110°C in 0.2 ml constant-boiling HCl, containing 0.1% phenol, in evacuated and sealed tubes. Analyses were performed on a Durrum D-500 amino acid analyzer. The colour value for *S*-carboxymethyl-cysteine was considered to be 92% that of aspartic acid.

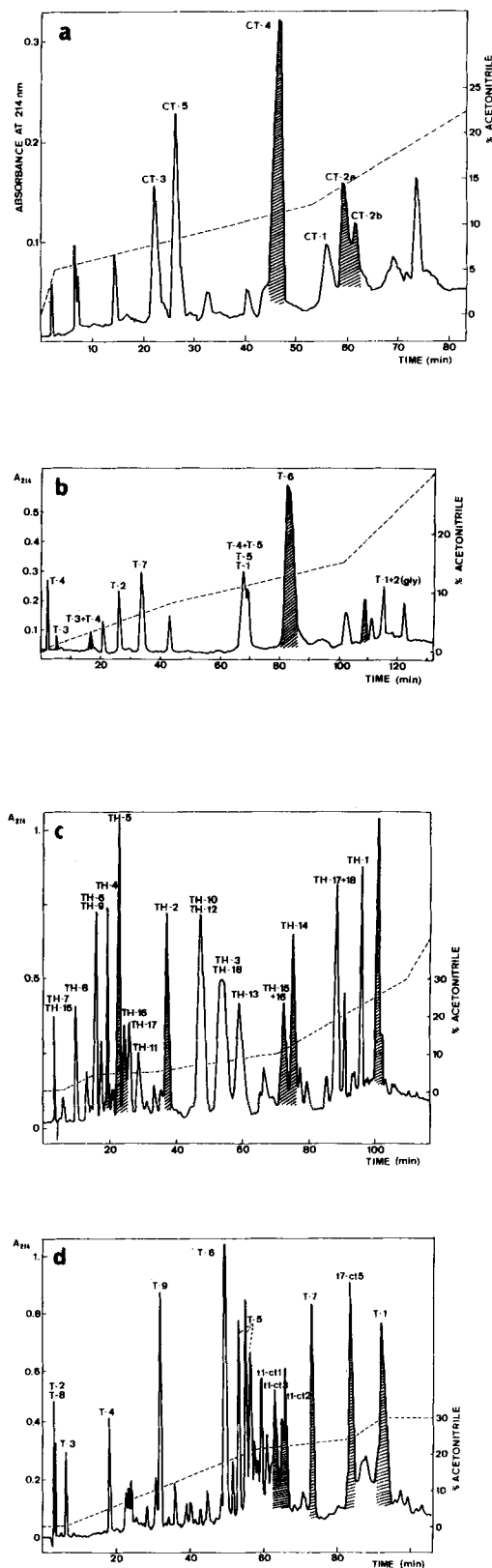


Fig. 2. Elution profiles of the reversed-phase chromatography on a Nucleosil 10 C_{18} column of different yellow mustard allergen hydrolysates. (a) Chymotryptic and (b) tryptic hydrolysates of the small chain; (c) thermolytic and (d) tryptic hydrolysates of the large chain. The peptides were eluted with an acetonitrile gradient in 0.1% (by vol.) trifluoroacetic acid. The flow rate was 1 ml/min. Fractions of 0.8 ml were collected. Dashed peaks corresponds to radioactive peptides

Sequence determination

Automatic Edman degradations were performed with a Beckman sequencer model 890C. 30 nmol of the uncleaved, reduced and carboxyamidomethylated large chain of the yellow mustard allergen, *Sin a* I, and about 15 nmol of a tryptic peptide were sequenced in the presence of 2–3 mg Polybrene, using the standard Beckman 0.1 M Quadrol program.

An Edman manual procedure for stepwise degradation [17, 18] was used for peptide sequence determination. The amount of each

peptide ranged over 10–40 nmol. The amino acids were resolved and identified using two different procedures: (a) reversed-phase HPLC [19] and (b) back-hydrolysis from the anilinothiazolinone derivatives, in constant-boiling HCl, containing 0.1% (mass/vol.) SnCl₂, for 6 h at 150 °C [20]. The presence of carboxyamido-methylcysteine was determined by measuring the radioactivity of a 20% (by vol.) aliquot of all the sequenced residues. Quantitative data from the Edman degradations are given in Table 6.

Table 1. Amino acid compositions of the chymotryptic peptides (CT-1 to CT-5) of the small chain of the *Sin a* I allergen. The number of residues (nearest integer) per peptide is given between parentheses. Tryptophan was identified by both absorption measures at 280 nm and manual Edman degradation.

	CT-1	CT-2a	CT-2b	CT-3	CT-4	CT-5
ASX						
THR						
SER						
GLX		1.3 (1)	1.4 (1)	3.1 (3)	2.0 (2)	2.6 (3)
PRO	1.6 (2)	1.2 (1)	1.2 (1)			2.2 (2)
GLY	1.1 (1)		1.3 (1)			1.3 (1)
ALA	1.1 (1)			1.0 (1)	1.0 (1)	2.0 (2)
Cys ^a		0.9 (1)	0.8 (1)		0.7 (1)	1.1 (1)
VAL						
MET						
ILE		0.9 (1)	0.9 (1)			0.8 (1)
LEU				0.9 (1)		0.9 (1)
TYR						
PHE	0.9 (1)	1.2 (1)	1.0 (1)	1.1 (1)		1.0 (1)
HIS		2.0 (2)	1.8 (2)			1.0 (1)
LYS		2.2 (2)	1.1 (1)		0.8 (1)	
ARG						
TRP						
Sequence position	1–5	6–14	6–14	15–20	21–26	27–39
Total residues	5	9	9	6	6	13

^aDetermined as carboxymethylcysteine

Table 2. Amino acid compositions of the tryptic peptides (T-1 to T-7) of the small chain of the *Sin a* I allergen. The number of residues (nearest integer) per peptide is given between parentheses. Tryptophan was identified by both absorption measures at 280 nm and manual Edman degradation.

	T-1	T-2	T-3	T-4	T-5	T-6	T-7
ASX							
THR							
SER							
GLX					4.1 (4)	2.2 (2)	2.6 (3)
PRO	2.2 (2)	0.8 (1)					2.0 (2)
GLY	1.0 (1)						1.3 (1)
ALA	0.9 (1)				0.9 (1)	1.2 (1)	1.9 (2)
Cys ^a			0.9 (1)			0.8 (1)	1.0 (1)
VAL							
MET							
ILE		1.0 (1)					0.9 (1)
LEU					0.9 (1)	1.0 (1)	
TYR							
PHE	1.0 (1)				0.9 (1)		
HIS					1.0 (1)	0.9 (1)	
LYS		1.0 (1)		1.0 (1)		1.0 (1)	
ARG	1.0 (1)		1.0 (1)		0.9 (1)		
TRP							
Sequence position	1–6	7–9	10–11	12	13–21	22–29	30–39
Total residues	6	3	2	1	9	8	10

^aDetermined as carboxymethylcysteine

Table 3. Amino acid compositions of the thermolytic peptides (TH-1 to TH-18) of the large chain of the yellow mustard allergen *Sin a* I.

The number of residues (nearest integer) per peptide is given between parentheses.

	TH-1	TH-2	TH-3	TH-4	TH-5	TH-6	TH-7	TH-8	TH-9	TH-10	TH-11	TH-12	TH-13	TH-14	TH-15	TH-16	TH-17	TH-18
ASX		1.0 (1)												1.1 (1)				
THR					1.0 (1)							1.0 (1)	1.0 (1)					1.0 (1)
SER							0.8 (1)				1.0 (1)				0.9 (1)			1.0 (1)
GLX	3.1 (3)	3.0 (3)	2.9 (3)					1.9 (2)	2.0 (2)	4.7 (5)		1.2 (1)	1.0 (1)	1.0 (1)				
PRO	3.6 (4)		1.4 (1)		0.9 (1)					1.0 (1)			1.0 (1)	1.0 (1)		1.0 (1)		1.8 (2)
GLY	1.1 (1)					1.1 (1)				1.8 (2)								1.2 (1)
ALA							2.1 (2)						1.0 (1)					
Cys ^a		1.7 (2)		1.0 (1)	1.0 (1)			1.0 (1)	0.9 (1)		0.7 (1)			1.0 (1)	1.0 (1)	1.0 (1)		
VAL														0.9 (1)	1.0 (1)	0.9 (1)		
MET																	1.0 (1)	
ILE																		
LEU	1.0 (1)	0.8 (1)	1.0 (1)	0.9 (1)		1.0 (1)				1.0 (1)	0.8 (1)	0.9 (1)	1.0 (1)					
TYR																		
PHE																	1.0 (1)	
HIS			1.0 (1)							1.0 (1)			1.1 (1)				1.8 (2)	
LYS						1.0 (1)	0.9 (1)	0.8 (1)					1.0 (1)					
ARG	1.0 (1)							1.0 (1)			1.2 (1)							
Sequence position	1–10	11–17	18–23	24–25	26–29	30–32	33–36	37–40	41–44	45–54	55–58	59–62	63–68	69–74	75–76	77–79	80–82	83–88
Total residues	10	7	6	2	4	3	4	4	4	10	4	4	5	6	2	3	3	6

^aDetermined as carboxymethylcysteine

Table 4. Amino acid compositions of the tryptic peptides (T-1 to T-9) of the large chain of the yellow mustard allergen, *Sin a* I. The number of residues (nearest integer) per peptide is given between parentheses.

	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9
ASX	1.3 (1)								
THR	1.1 (1)					1.9 (2)			0.8 (1)
SER		1.0 (1)			1.2 (1)		0.9 (1)		0.9 (1)
GLX	9.0 (9)			2.2 (2)	6.9 (7)	0.9 (1)	1.2 (1)		
PRO	5.6 (6)				0.9 (1)	1.0 (1)	2.1 (2)		2.1 (2)
GLY	1.3 (1)	1.0 (1)			2.2 (2)				1.0 (1)
ALA		1.0 (1)	1.0 (1)						
Cys ^a	3.8 (4)					1.1 (1)			
VAL	1.1 (1)		1.0 (1)	1.0 (1)	1.2 (1)		2.0 (2)		
MET							2.8 (3)		
ILE					0.9 (1)	1.0 (1)			1.0 (1)
LEU	4.6 (5)				1.1 (1)	1.1 (1)	1.0 (1)		
TYR						0.9 (1)			
PHE							1.0 (1)		
HIS	1.0 (1)				1.0 (1)	1.1 (1)			
LYS	1.0 (1)	1.1 (1)	1.0 (1)			1.0 (1)	0.9 (1)	1.0 (1)	
ARG	1.0 (1)			1.0 (1)	1.0 (1)				
Sequence position	1–31	32–35	36–38	39–42	43–58	59–68	69–81	82	83–88
Total residues	31	4	3	4	16	10	13	1	6

^aDetermined as carboxymethylcysteine

Table 5. Amino acid composition of the major allergen of yellow mustard seeds, *Sin a* I, and its polypeptide chains.

	NATIVE PROTEIN		LARGE CHAIN		SMALL CHAIN	
	Acid hydrolysis	Sequence	Acid hydrolysis	Sequence	Acid hydrolysis	Sequence
ASX	2	2	2	2	–	–
THR	4	4	4	4	–	–
SER	6	7	3–4	4	3	3
GLX	27–28	28	19	20	8–9	8
PRO	15–16	16	10–11	12	4–5	4
GLY	8	8 ^a	5	3	3	3 ^a
ALA	7	7	4	3	4	4
Cys ^a	8 ^a	8	6 ^a	6	2 ^a	2
VAL	6	7	6	7	–	–
MET	2	2	1	1	1	1
ILE	4	4	3	3	1	1
LEU	8	9	6	7	2	2
TYR	1	1	1	1	–	–
PHE	3	3	1	1	2	2
HIS	4–5	5	3	3	2	2
LYS	7–8	9	5	6	2–3	3
ARG	7	5 ^a	4	3	3	2 ^a
TRP	1 ^b	1	–	–	1 ^b	1
Total residues	120–124	127	83–85	88	37–40	39

^aDetermined as cysteic acid after 24 h hydrolysis

^bDetermined spectrophotometrically by using the method of Beaven and Holiday (24)

^cHalf residue is considered as a consequence of the heterogeneity found at the sixth position in the small polypeptide chain of the allergen.

Table 6.- Quantitative data concerning the Edman degradation experiments.

SMALL CHAIN			
Total chain (30)	CT-2a (10)	CT-2b (10)	CT-3 (10)
1. Pro 13.8	1. Arg 2.8	1. Gly 2.1	1. Gln 4.4
2. Ala 7.7	2. Ile 2.9	2. Ile 2.0	2. Gln 3.8
3. Gly 12.0	3. Pro 2.5	3. Pro 1.3	3. Ala 3.5
4. Pro 10.4	4. Lys 0.8		4. Gln 2.8
5. Phe 10.0	5. Cys * 1.0	T-1 (7)	5. His 1.8
6. Arg 2.6	6. Arg 1.0	1. Pro 3.6	6. Leu 1.2
Gly 4.2	7. Lys 0.8	2. Ala 1.9	
7. Ile 9.1	8. Glu 0.4	3. Gly 2.5	T-2 (10)
8. Pro 6.5	9. Phe 0.6	4. Pro 2.2	1. Ile 6.3
9. Lys 0.7		5. Phe 1.6	2. Pro 4.1
10. Cys * 3.6	CT-5 (10)	6. Arg 2.9	3. Lys 3.2
	1. Leu 6.4		
CT-4 (10)	2. His 3.4	T-3 (5)	T-6 (15)
1. Arg 4.0	3. Lys 1.4	1. Cys * 2.5	1. Ala 8.2
2. Ala 5.1	4. Gln 2.2	2. Arg 3.9	2. Cys * 6.0
3. Cys * 1.7	5. Ala 2.5		3. Gln 6.5
4. Gln 1.8	6. Met 1.1	T-5 (5)	4. Gln 7.2
5. Gln 1.7	7. Gln 1.3	1. Glu 2.2	5. Trp 4.2
CN-1 (50-P) (20)	8. Ser 0.3	2. Phe 3.0	6. Leu 5.3
1. Ser 5.2		3. Gln 1.6	7. His 3.1
2. Gly 13.6			
3. Ser 3.1			
4. Gly 10.2			
5. Pro 10.4			
6. Ser 1.4			
LARGE CHAIN			
Total chain (30)	TH-2 (35)	TH-3 (25)	TH-4 (40)
1. Pro 8.5	1. Leu 29.1	1. Leu 16.9	1. Leu 30.9
2. Gln 8.3	2. Gln 25.3	2. His 16.8	2. Cys * 25.0
3. Gly 8.3	3. Gln 20.1	3. Gln 14.6	
4. Pro 6.8	4. Cys * 15.3	4. Glu 18.0	TH-7 (40)
5. Gln 6.5	5. Cys * 13.5	5. Glu 13.8	1. Ala 34.5
6. Gln 6.3	6. Asn 15.4	6. Pro 6.8	2. Lys 25.7
7. Arg 5.4	7. Glu 13.6		3. Ser 5.1
8. Pro 3.3		TH-6 (35)	4. Ala 12.4
9. Pro 3.2	TH-5 (40)	1. Leu 29.7	
10. Leu 2.8	1. Val 34.7	2. Lys 28.0	TH-10 (25)
11. Leu 2.7	2. Cys * 14.7	3. Gly 15.7	1. Leu 18.8
12. Gln 2.3	3. Pro 29.5		2. Glu 14.3
13. Gln 2.2	4. Thr ** 28.6	TH-9 (40)	3. Gln 12.6
14. Cys * 1.8		1. Val 33.7	4. Gln 9.4
15. Cys * 1.7	TH-8 (40)	2. Arg 25.7	5. Gly 3.6
16. Asn 2.0	1. Val 38.0	3. Gln 23.0	6. Gln 6.9
17. Glu 1.6	2. Lys 18.5	4. Gln 20.3	7. Gln 6.3
18. Leu 2.1	3. Gln 16.7		8. Gly 6.6
19. His 1.3	4. Gln 18.3	TH-12 (35)	9. Pro 4.3
TH-14 (40)		1. Ile 28.8	10. His 5.6
1. Val 32.1	TH-11 (40)	2. Tyr 11.2	
2. Cys * 12.5	1. Val 32.7	3. Gln 12.1	TH-13 (25)
3. Asn 21.3	2. Ile 24.3	4. Thr ** 10.3	1. Ala 17.8
4. Ile 8.4	3. Ser 7.0		2. Thr ** 5.2
5. Pro 14.6	4. Arg 22.0	TH-16 (35)	
6. Gln 16.0		1. Val 25.0	TH-17 (25)
TH-18 (15)	TH-15 (40)	2. Cys * 12.9	1. Phe 20.0
1. Thr 3.3	1. Val 35.3	3. Pro 20.8	2. Lys 22.1
2. Met 5.9	2. Ser ** 39.7		3. Lys 17.3
3. Pro 7.1	T-2 (25)	T-3 (25)	T-4 (25)
4. Gly 6.9	1. Gly 21.0	1. Ala 17.5	1. Gln 21.4
5. Pro 10.7	2. Ala 17.6	2. Val 19.2	2. Gln 18.2
6. Ser ** 4.9	3. Ser 3.2	3. Lys 4.3	3. Val 17.9
T-5 (15)	4. Lys 25.2	T-7 (15)	4. Arg 15.6
1. Gln 7.3		1. Val 11.5	
2. Gln 4.7	T-6 (15)	2. Cys * 3.7	T-9 (15)
3. Leu 5.4	1. Ile 7.4	3. Asn 5.6	1. Thr 2.3
4. Glu 6.0	2. Tyr 6.9	4. Ile 6.4	2. Met 11.7
5. Gln 2.5	3. Gln 5.1	5. Pro 4.7	3. Pro 6.3
6. Gln 2.8	4. Thr 1.3	6. Gln 2.7	4. Gly 6.6
7. Gly 3.3	5. Ala 2.6	7. Val 3.4	5. Pro 4.2
8. Gln 1.7	6. Thr 0.7	8. n.d. 3.6	6. Ser ** 5.3
9. Gln 2.4	7. His 1.4	9. Val 3.6	
10. Gly 2.1	8. Leu 0.6	10. Cys * 1.5	CT-6 (12)
11. Pro 1.8	9. Pro 0.8		1. Lys 9.0
12. His 1.5	10. Lys 0.4	CT-4 (10)	2. Lys 8.5
13. Val 2.2		1. Val 6.0	3. Thr 2.2
		2. Ile 4.8	4. Met 7.8
		3. Ser 1.2	5. Pro 4.9
		4. Arg 3.3	6. Gly 8.1
		5. Ile 3.6	7. Pro 7.1
		6. Tyr 2.7	8. Ser ** 8.0

(*) Determined from radioactive counting.

(**) Determined from the aqueous phase.

n.d. Not determined.

Numbers between parenthesis correspond to total nmol of each peptide subjected to Edman degradation.

Values are expressed in nmol of each residue.